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 and Reyn van Ewijk

October 2017, updated June 2021

Discussion paper number 2110

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

Contact Details:

Reyn van Ewijk
Chair of Statistics and Econometrics
Johannes Gutenberg-University Mainz
Jakob-Welder-Weg 4
55128 Mainz
Germany

vanewijk@uni-mainz.de

No "Honeymoon Phase"

Whose health benefits from retirement and when

Birgit Leimer *

Reyn van Ewijk *

Abstract

We use a fixed effects instrumental variable approach to investigate whether retirement affects health only temporarily during a honeymoon phase or if health effects materialize after retirement and remain. The normal and early retirement age thresholds are used as instruments. 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), we find that retiring both at the normal and early retirement eligibility ages significantly improves all health aspects, including the objective measure grip strength. Other than hypothesized, results do not show a health boost during the honeymoon phase. Instead, individuals, especially blue-collar workers, go through an adjustment period after retiring, 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

* Faculty of Law and Economics, Johannes Gutenberg-University Mainz, Germany. Correspondence: Reyn van Ewijk, Chair of Statistics and Econometrics, Johannes Gutenberg-University Mainz, Jakob-Welder-Weg 4, 55128 Mainz, Germany. Email: vanewijk@uni-mainz.de.

1. Introduction

Even though life expectancy has doubled over the last century (World Bank 2016b), retirement age thresholds have decreased since Otto von Bismarck introduced a retirement age of 70 in 1889 (Herbay 2014). This has led 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 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 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, such as unemployment or disability leave, 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 medical expenditure. Following this argument, postponing retirement might be beneficial for retirees. It is therefore unclear, whether health is preserved, unchanged or harmed by retirement.

Providing causal evidence on the impact of retirement on health is not straightforward, as reverse causality and omitted variable bias may exist. On the one hand, poor health and health shocks influence a person's decision to retire (Dwyer and Mitchell 1999). On the other hand, an individual's observable and unobservable characteristics may drive the retirement decision and influence the health status. Several studies have attempted to account for these endogeneity concerns by using stratified samples or instrumental variables. However, no definite conclusions can be drawn as opposing results have been presented.

The inconclusive results may be explained by heterogeneity in effects across the various health dimensions that have been considered, heterogeneity in effects by background characteristics including the type of job previously held, as well as heterogeneity in effects across the conditions under which the retirement occurs (e.g. retirement due to an unexpected offer to retire early may have different health effects than a retirement that is due to reaching the normal retirement age). Finally, the heterogeneity of the retirement effect needs to be considered not only across different groups, but also over time. 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 temporary idealistic state immediately after retiring. This is expected to have a health improving effect, especially on measures such as self-assessed health and mental health. Previous research that incorporated this time path into the effect of retirement either only compared effects between two pre-specified periods that were often dictated by data availability (usually: effects within about two years of retirement vs effects after about four years, which corresponds to the data for when having at most three waves of a biannual survey). Or it was implicitly assumed that there would be a one-off jump in health, followed by a linear change in health. We instead investigate the length of a potential honeymoon phase, taking a flexible approach regarding the time path of the effect of retirement.

To give an encompassing analysis of the health effects of retirement, we furthermore 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, as mentioned, split into the honeymoon phase and later retirement phase.

We 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 to 80-year-old individuals who have been employed or retired in all waves, living in Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden and Switzerland.

Results of our analysis show that retiring leads to subjective and objective health improvements. Noteworthy is the result that retirees have significantly higher maximum grip strength, showing that health benefits are not solely found in subjective health measures. The health preserving effect remains when separately controlling for a honeymoon effect. Opposite to our hypothesis, there is no health boost during the honeymoon phase. Instead, there is evidence of the opposite occurring - retirees, especially blue-collar workers, first experience significantly worse health upon retiring. We further find that both genders experience health improvements. Women experience greater improvements in terms of depression and grip strength. 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.

The paper proceeds as follows. Section 2 presents the theoretical impact of retirement on health,gives an overview of the current literature and explains our contribution to the 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, which also concludes this paper.

2. Background

2.1. 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) is often used when discussing the theoretical effect of retirement on health. It views health to be both an investment and a consumption good. Investing in health, through health-promoting activities, 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.

Further theories support retirement having either a beneficial or detrimental role in health. As Kim and Moen (2002) point out, the role change due to retirement could mean a feeling of identity loss for those individuals whose work was a central part of their identity. On the other hand, it could also reduce role strain and overload, resulting in fewer depressive symptoms and better overall health. 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, may lose their meaning in life with retirement. Furthermore, social interactions, which have been linked to better health outcomes (Petrou and Kupek 2008), may also change drastically with retirement. There is evidence that especially women benefit from the additional free time after retiring to spend with friends and family (Thomas 2011). This could lead to 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 of 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.

2.2. Literature Review and Contribution

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 and Mitchell 1999). To fully account for the endogeneity caused by reverse causality, it has become widely accepted to use an instrumental variables (IV) approach.

Several studies used offers of early retirement as an instrument (Coe & Lindeboom, 2008; Coe et al., 2012; Hallberg et al., 2015). 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, this might limit external validity as in most cases behavioral adjustments can be considered part of the treatment, while moreover individuals who are offered early retirement packages tend not to be representative of all workers. Insler (2014) used self-reported probabilities to work past the ages 62 and 65 as an instrument. These were taken long before the actual retirement age, and adjusted for a number of observables so that these probabilities may be thought of as exogenous. Another line of research utilized changes to the expected remaining working time due to changes to pension schemes (Atalay et al., 2019; Bertoni et al., 2018; Bloemen et al., 2017; Hagen, 2017; Hernaes et al., 2013; Shai, 2018). While these studies are very relevant when considering the health effects of retirement, having to continue working longer may lead to a rather specific local average treatment effect, especially if the change is unexpected or applied to only some workers (defined by type of occupation or cohort), but not to their peers with different occupations or born in slightly different years (De Grip et al., 2012).

By far the most common instruments are the normal and early retirement eligibility ages (NRA and ERA). Reaching these eligibility ages increases the probability of an individual to retire, without having a separate effect on health. Early studies used cross-sectional data and found mixed results (Bound & Waidmann, 2007; Coe and Zamarro, 2011; Rohwedder and Willis, 2010). To better account for endogeneity issues and to obtain more precise results, further studies used panel data to identify the direction of the effect (e.g. Behncke, 2012; Bertoni et al., 2018; Bonsang et al., 2012; Gorry et al., 2018; Heller-Sahlgren, 2017; Kolodziej and García-Gómez, 2019; Mazzona and Peracchi, 2012 and 2017; Neuman, 2008). There are still a few potential pitfalls to consider for such studies. One is a possible violation of the exogeneity assumption when using the NRA as an instrument in US datasets (Bonsang et al., 2012; Gorry et al., 2018; Rohwedder and 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).To circumvent this issue, Neuman (2008) excluded the NRA in his set of instruments. The other issue is a failure to control for unobserved individual heterogeneity. This methodological choice impacts the identified effect, as discussed in Nishimura et al. (2018). Those studies that do control for individual level fixed effects alongside their IV strategy still found opposing effects.

By now, a wide array of outcomes has been studied with usually differing results. One of the few outcomes in which results were quite consistent across studies is self-assessed health (SAH). Most studies looking at SAH found that people perceived their own health to be better after retirement (Coe and Zamarro, 2011; Eibich, 2015; Gorry, 2018; Grøtting & Lillebø, 2020; Neuman, 2008). Mazzonna and Peracchi (2017) found that the improvements in SAH only occurred among those who had been working in physically demanding occupations.

This finding does not clearly translate into objective health outcomes. For mortality, most studies reported no effects (Grøtting & Lillebø, 2020; Hagen, 2017; Hernaes et al., 2013; Nielsen, 2019) while Fitzpatrick & Moore (2018) found that retirement increases mortality rates and Bloemen et al. (2017) found decreased mortality rates. Also results using other health outcomes such as the number of chronic conditions, disease diagnosis, or limitations in (instrumental) activities in daily living, have led to mixed results without an apparent overall trend.

Also the results for mental health (usually measured as depressive symptoms) have been mixed, with some studies finding positive effects (Eibich, 2015; Gorry, 2018; Kolodziej and García-Gómez, 2019), and others no effects (Coe and Zamarro, 2011; Neuman, 2008). Heller-Sahlgren (2017) found that retirement has no short-term mental health effect, but a large negative longer-term effect. And Mazzonna and Peracchi (2017) reported adverse effects for a large part of the population, but improvements for those who were previously in physically demanding occupations.

Several studies found that cognitive abilities declined with retirement (Atalay et al., 2019; Bonsang et al., 2012; Mazzonna and Peracchi, 2012; Rohwedder and Willis, 2010), while Coe et al. (2012) found no effects and Celidoni et al. (2017) found that retiring at the early retirement age leads to a protective effect while retiring at the normal retirement age leads to a long-term negative effect.

Several studies looked at health care utilization and – once more – found varying results (Eibich, 2015; Hagen, 2017; Lucifora & Vigani, 2017; Nielsen, 2019; Zhang et al., 2019). These results may not be purely interpretable as health effects, as retirement may also change the utilization of many types of health care, such as doctor visits, due to a decrease in the opportunity cost of time. This is much less the case for hospitalizations. Grøtting & Lillebø (2020) found that retirement led to fewer acute hospitalizations, while Nielsen (2019) found that retiring at the early retirement age led to fewer hospitalizations while there were no effects for retiring at the normal retirement age.

Finally, several studies considered health behaviors and lifestyle which can be an important channel for effects of retirement on health. They particularly focused on whether retirees are more likely to adopt a sedentary lifestyle or whether they become more active. Celidoni (2017) and Eibich (2015) found that retirement leads to increases in activity levels, while Fe and Hollingsworh (2016) in contrast found that retirement leads to a more sedentary lifestyle. Godard (2016) found increases in BMI and obesity rates, but only for males who had been working in strenuous jobs or who had already been at risk of obesity.

Given this considerable heterogeneity in previously reported results, it is not easy to come to a single, overarching conclusion regarding the effects of retirement. This heterogeneity seems to be due to the following causes. First, to different ways in dealing with endogeneity. However, only considering those studies that do a very thorough job at taking out endogeneity concerns does not change the picture that effects vary strongly. Second, the effect of retirement may differ between the various types of health outcomes that have been studied. Again, it will be clear that this cannot explain the full heterogeneity in effects summarized above. Third, the effect may be inherently diverse, with heterogeneity according to pre-existing characteristics, or to the conditions under which the retirement occurs.

Regarding these pre-existing characteristics, studies differ in their studied age spans as well as in their decision whether to include individuals who are unemployed, homemakers and permanently disabled. It is unclear how these choices affect the results. In terms of heterogeneity according to pre-existing characteristics, 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 and 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) and Kolodziej and García-Gómez (2019) explored the heterogeneity among occupational groups and found that blue-collar workers experienced a significant health improvement. Mazzonna and 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 retirement has a negative effect on health, except amongst individuals in particularly physically burdensome jobs.

Regarding heterogeneity in effects according to the conditions under which the retirement occurs, it is particularly relevant that studies differed in the local average treatment effects (LATE) that they studied. Studies that utilize changes to retirement age find effects that may be partially driven by comparisons of one's own situation to that of unaffected peers who happen to be slightly older or in a different occupational group. This LATE may be different from effects of a retirement that is induced by simply reaching the early or normal retirement age. And the LATEs associated with the latter two can also differ between each other, as should be clear from the literature review above.

The fourth cause of heterogeneity in reported effects is that the effect of retirement may follow a certain time path that is not fully considered in most studies. Many studies essentially compared retirees with non-retired people. Some studies have attempt to consider the dynamic effect retirement has on health or whether identified effects are immediate and/or temporary. It is thought that the effect of retirement may be separable into several phases (Atchley 1980). Several studies have picked up the idea of one particular phase immediately after retiring, called the honeymoon phase, which is believed to be a euphoric period, in which retirees enjoy the new-found freedom, time and space. The idea being that the positive aspects of retirement outweigh potential negative aspects. Especially self-assessed health measures and mental health could improve drastically through this feeling of euphoria. For policy makers it will be important to understand if health improvements are temporary or more sustained over a longer time frame.

Several studies split the effect of retirement into a short and a long-term effect, often by analyzing the effect retiring had after one wave and after two waves of a longitudinal survey (Coe & Lindeboom, 2008; Gorry et al., 2018; Heller-Sahlgren, 2017; Insler, 2014). The limitation of this is that it imposes a certain length of the honeymoon phase and that longer-run effects could not be considered. Other studies explicitly modeled the time path of the effect of retirement. These include Bonsang et al.

(2012), Fe & Hollingsworth (2016), Lucifora & Vigani (2017) and Mazzonna & Peracchi (2017). These studies' models all assumed that health may exhibit a one-off drop or jump at the moment of retirement and that the retirement effect afterwards develops linearly over time.[1] This makes a strong assumption on the health path after retirement (a sudden shock followed by a linear change) and may lead to wrong predictions for health in the long-run as there is no bound to the linear change in health after the initial drop or jump.

We deviate from this modeling assumption by explicitly defining a honeymoon phase of a certain post-retirement period during which the health effect may differ from that during later periods. We investigate different potential lengths of this phase, as there is no theoretical justification for claiming that the honeymoon phase should have a specific length. This is rather something that should be empirically investigated. Our analyses thus allow for a very flexible approach regarding the time path of the effect of retirement.

Our study's first contribution lies in this investigation of the honeymoon effect. A second contribution is that we are the first in this domain to use five waves of a longitudinal dataset on elderly. Following the same set of individuals over up to about 11 years is helpful when investigating the time path of retirement effects, as well as for investigating the robustness of results against alternative age specifications in our individual fixed effects models, including up to a quartic polynomial in age.

Finally, we study a total of ten countries, we study a wide range of health outcomes, focusing on both objective and subjective health measures in both the domains of physical and mental health, and we study a range of relevant potential effect heterogeneities (by sex, white vs blue collar worker and type of job (physical, time pressure, freedom in job)). We furthermore investigate the differences in the health effects of retiring at the NRA and at the ERA. For none of these separately we are the first, but all taken together provides for an encompassing overview of the effects of retirement on health.

3 Data

This paper uses data from SHARE Waves 1, 2, 4, 5 and 6 (Börsch-Supan 2013). 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. Similarly, the seventh wave was – for most individuals – a retrospective survey, and is therefore not included in this paper.

---

[1] In addition, two papers (Bonsang & Klein (2012) and Kesavayuth et al.(2020)) modeled the time path around retirement more flexibly using a set of dummies indicating whether someone has retired <1 year ago, 1-2 years age, etc. These papers however do not study health outcomes but measures of life satisfaction.

3.1. Sample Selection and Retirement Definition

Our sample consists of all countries which were surveyed in all the five waves that we used. These are Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden and Switzerland. We restricted the sample to individuals who are observed at least twice and who are between the ages of 50 and 80.

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	58.0	78.6%	55.1%	9.2	2.0	6 674
Belgium	64.1	60.0	62.3%	44.1%	12.3	2.1	6 927
Denmark	63.3	62.8	47.8%	48.7%	13.5	2.3	5 935
France	65.2	59.8	68.5%	50.5%	11.2	2.2	6 814
Germany	65.3	61.2	62.9%	47.7%	12.5	2.0	6 425
Italy	65.8	57.7	74.3%	37.5%	8.8	1.9	5 160
Netherlands	64.2	62.1	56.4%	38.0%	11.8	2.3	4 129
Spain	66.8	61.8	68.3%	29.5%	8.5	2.2	4 822
Sweden	66.5	63	63.0%	53.8%	11.6	2.4	8 186
Switzerland	64.3	62.5	51.6%	47.7%	9.1	2.0	5 287
Total	65.2	60.8	63.7%	46.4%	11	2.1	60 359

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 retired, while those declaring themselves to be employed are considered working. To measure the effect of transitioning into retirement from employment, individuals who ever report any other job status are dropped from the analysis. Furthermore, as civil servants and self-employed individuals often face different retirement eligibility rules, individuals who reported being either are excluded in the main analysis.

There is a total of 60 359 person-observations for 21 212 individuals in the main analysis. Around 19% of these individuals transition from work to retirement in the period of observation. The percentage of retired individuals differs between countries, as is shown in Table 1. This is in part explained by the different retirement eligibility ages as well as the different attitudes toward retirement in the different countries.

3.2. Retirement Eligibility Ages of the Sample

The SHARE dataset is supplemented with the relevant normal and early retirement age thresholds after which an individual is eligible to receive pension benefits. The eligibility ages are gender, cohort, year and country specific. Table 2 gives an overview of the most common eligibility ages in the interview years. 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 63 for women.[2] Over the time span of the interviews, retirement ages have increased and some early retirement schemes have already been abolished.

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.

3.3. Health Measures

Several previous research papers have focused on specific health aspects when investigating the relationship between retirement and health. In this paper, six health measures are used separately as the outcome variable to gain a comprehensive understanding. The health measures include subjective and objective measures as well as physical and mental health aspects. Policy conclusions should be based on understanding how health overall is affected by retirement instead of only considering the health effect in one realm.

The first health measure, capturing general well-being, 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).

[2] 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).

Following convention, an indicator variable is generated which is equal to 1 if a person reports being in very good or excellent health and 0 otherwise. SAH has been shown to be an independent predictor of mortality, particularly among the elderly (see for example Idler and Benyamini (1998)).

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 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-recall tests are summed up, giving a maximum score of 20. Total word recall is used instead of other available cognitive ability variables, 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. One indicator variable is generated that is equal to 1, if a 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). It has been shown that grip strength is a good, independent predictor of mortality (Ambrasat et al. (2011), Hank et al. (2009) and Leong et al. (2015) among others). Grip strength is the only truly objective health measure available.

Two 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 time is less restricted during retirement, the opportunity 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. We do not include a health index in which several health variables are used to measure a person’s general health, as this will not reveal potentially heterogenous effects of retirement.

4. Methods

The challenge in identifying a causal relationship between retirement and health is twofold. On the one hand, many individual characteristics influence both a person’s decision to retire and their health. As a result, the health of retirees and workers cannot be simply compared. While some of these characteristics are observable – such as age, gender, industry, etc. – others, such as preferences or past experience, are not. This could lead to biased results. The second source of endogeneity is reverse causality. It has been shown that the retirement decision is in part driven by poor health (Dwyer and Mitchell 1999).

4.1. Fixed Effects Instrumental Variable Model

4.1.1. Binary Retirement Decision

The endogeneity concerns are dealt with using two approaches. On the one hand, an individual level fixed effects (FE) approach is used to control for unobserved time-invariant heterogeneity:

$$Y_{it} = \beta_1 R_{it} + \alpha X_{it} + \mu_i + \tau_t + u_{it} \quad (2.1)$$

where i is the individual, t the survey period, Y_{it} a health measure, X_{it} a vector of control variables and R_{it} an indicator equal to 1, if a person is retired. μ_i are individual and τ_t wave fixed effects. Age and age squared are used as control variables. We will test different age trend specifications to ensure the robustness of our results. Using an FE-approach ensures that time-invariant confounding factors, such as gender, years of education, number of children, and country of residence, as well as genes or preferences, are controlled for.

This model may still suffer from endogeneity, as individual fixed effects will not remove reverse causality. Following the conventional approach, we exploit the fact that the 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 NRA and ERA thresholds 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. A two stage least square (2SLS) estimation procedure is used. The following first stage regression is estimated:

$$R_{it} = \delta Z_{igt} + \varphi X_{it} + \mu_i + \tau_t + \epsilon_{igt} \quad (2.2)$$

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, $\hat{R}_{it}$, are used to estimate:

$$Y_{it} = \beta \hat{R}_{it} + \rho X_{it} + \varsigma_i + \theta_t + u_{igt} \quad (2.3)$$

4.1.2. Honeymoon Phase

In a second step, the effect of retirement is separated into honeymoon and retirement phases. A person's retirement date is used to calculate her exact time since retirement. It is a priori not clear how long a honeymoon phase can last, therefore lengths between 6 and 36 will be considered (in six-month steps). We now have two endogenous variables: H_{it} is an indicator for being in the honeymoon phase and R_{it} is an indicator for being retired for longer than the honeymoon phase. These endogenous variables are instrumented using a set of gender-specific indicator variables, which are equal to 1 if an individual is within a certain time period after the ERA or NRA and indicators which are equal to 1 if an individual has been beyond the ERA or NRA for longer than this certain time period. The first set of instruments is thus defined as $I[NRA_{igt} \leq age_{igt} < NRA_{igt} + 6\ (to\ 36)\ months]$ and $I[ERA_{igt} \leq age_{igt} < ERA_{igt} + 6\ (to\ 36)\ months]$. And the second set of instruments is given by: $I[\,age_{igt} \geq NRA_{igt} + 6\ (to\ 36)\ months]$ and $I[age_{igt} \geq ERA_{igt} + 6\ (to\ 36)\ months]$. Including the two distinct phases leads to the following first stage regressions in which Z_{igt} includes all instruments:

$$R_{it} = \delta\, Z_{igt} + \varphi X_{it} + \mu_i + \tau_t + \epsilon_{igt} \qquad (2.4)$$

$$H_{it} = \eta\, Z_{igt} + \upsilon X_{it} + \chi_i + \psi_t + o_{igt} \qquad (2.5)$$

The second stage estimation becomes:

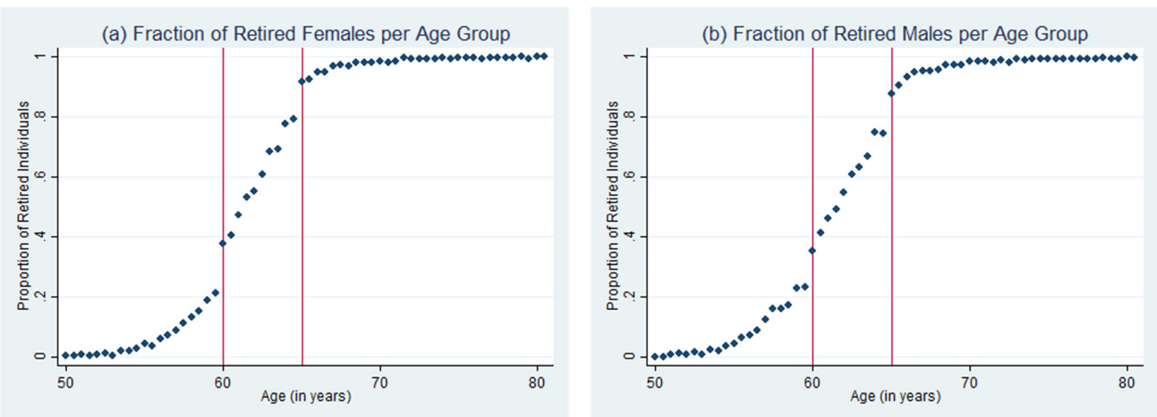
$$Y_{it} = \beta\, \hat{R}_{it} + \beta_2 \hat{H}_{it} + \rho X_{it} + \varsigma_i + \theta_t + u_{it} \qquad (2.6)$$

4.1.3. 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 1 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. There is a jump in the fraction of retired individuals of around 15-20 percentage 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. Analogous to the jumps in the probability to retire, there are jumps in the probability to be in the honeymoon phases.

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 a lower bound of the true effect. Cross country variation in retirement ages allows for an abstraction from this potential effect.

Figure 1: Proportion of Retirees Per Age in 6-Months Bins



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

However, several threats to the exogeneity assumption remain. Some studies argue that the NRA and ERA are known ahead of time, thereby causing an individual to alter behavior prior to retirement. Knowing there is only a limited time left in their job, individuals may be less stressed by their job. This 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. As a robustness check, this additional pre-retirement phase will be controlled for to ensure the validity of the results.

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 advanced 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 in the subsequent years will be excluded.

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 reaching a threshold age makes some persons more likely to retire and others 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 is violated.

4.2. Heterogeneity

Health trajectories differ between males and females, so heterogeneities by gender will be investigated. The effect of retirement will then be broken down further by job characteristics, including white- and blue-collar workers.[3] Table 3 shows that blue-collar workers, on average, experience greater health issues among retired and nonretired individuals. If blue-collar workers benefit more than white-collar workers, an average treatment effect may be insignificant due to the larger sample size of white-collar workers (there are over twice as many white-, than blue-collar workers in the sample). The effect is further differentiated using questions about the characteristics of the job to see if those who feel that their job is physically straining, who feel time pressure for a large workload or feel like they have no freedom in their work, benefit more from retirement. These questions ask the respondent to strongly agree, agree, disagree, or 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 to reduce endogeneity.

[3] Individuals are categorized as blue- or white-collar worker using the ISCO-88 categorization given in the first interview.

Table 3: Average Health Measures - Retired vs Working Individuals

	All Workers		White Collar		Blue Collar	
	Working	Retired	Working	Retired	Working	Retired
Age	57.0	69.8	57.0	69.6	56.8	70.3
Very Good / Excellent SAH	47.4%	27.7%	51.2%	64.0%	35.2%	18.9%
Not Depressed	83.6%	79.2%	83.8%	81.1%	82.8%	75.7%
Total Words Recalled	10.9	8.9	11.4	9.7	9.4	7.5
No (I)ADL Limitations	94.3%	86.1%	94.7%	85.4%	93.1%	79.1%
No Mobility Limitations	74.4%	52.9%	75.4%	55.9%	71.2%	47.6%
Maximum Grip Strength	39.7	34.8	38.6	34.3	43.2	35.5

5. Results

5.1. Retirement Effect on Health Outcomes

A negative association between retirement and health is suggested by Table 3. It shows that on average, workers are in better health. This negative correlation is confirmed by the pooled OLS (POLS) regression results shown in column (1) of Table 4. Retirement is associated with significantly worse health in five out of the six health outcomes. The sign of the association switches for five out of the six health measures once individual level fixed effects are controlled for (see column (2) of Table 4). Even though reverse causality is not yet accounted for, the negative association between retirement and health is no longer present, indicating that unobserved individual heterogeneity is driving a large part of the negative association.

To remove the bias due to reverse causality, Table 4 shows the results when using NRA (column (3)), ERA (column (4)) and both (column (5)) to instrument for the retirement decision. Instrument validity is checked using the Kleibergen-Paap rk Wald F statistic as well as the test for over- and under-identification. 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 and Yogo (2005) 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 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 (see Table 5). Reaching the normal retirement age increases the probability to retire by around 38% for women and 27% for men, while reaching the early retirement age threshold increases the probability by around 20% for women and 22% for men.

Table 4: Binary Effect of Retirement on Health of Entire Population

	POLS	FE	Instrumental Variable Approach		
	(1)	(2)	(3)	(4)	(5)
Very Good / Excellent Health					
retired	-0.050***	0.025***	0.086***	0.097***	0.088***
	(0.008)	(0.009)	(0.023)	(0.036)	(0.021)
KP F-stat			[1003]	[484]	[833]
N	58 954	60 344	59 813	59 813	59 813
Not Depressed					
retired	-0.003	0.018**	0.033*	0.059**	0.039**
	(0.006)	(0.007)	(0.020)	(0.030)	(0.017)
KP F-stat			[993]	[485]	[829]
N	58 079	59 401	58 549	58 549	58 549
Total Words Recalled					
retired	-0.136***	0.050	0.101	0.046	0.079
	(0.052)	(0.057)	(0.149)	(0.232)	(0.130)
KP F-stat			[959]	[480]	[812]
N	57 430	58 746	57 567	57 567	57 567
No (I)ADL Limitations					
retired	-0.011**	0.023***	0.058***	0.023	0.048***
	(0.005)	(0.006)	(0.016)	(0.022)	(0.014)
KP F-stat			[1003]	[484]	[833]
N	58 955	60 347	59 815	59 815	59 815
No Mobility Limitations					
retired	-0.036***	0.007	0.057**	0.074**	0.062***
	(0.008)	(0.009)	(0.024)	(0.035)	(0.020)
KP F-stat			[1001]	[484]	[832]
N	58 958	60 348	59 817	59 817	59 817
Maximum Grip Strength					
retired	-0.290**	-0.113	0.800***	0.135	0.560**
	(0.133)	(0.111)	(0.275)	(0.457)	(0.251)
KP F-stat			[938]	[471]	[797]
N	55 430	56 670	54 898	54 898	54 898
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 reported in parentheses. Kleibergen-Paap (KP) rk Wald F-statistics are reported in brackets. POLS stands for a pooled ordinary least square regression. FE is an individual-level fixed effects regression. POLS regressions control for age, age squared, female, number of children, years of education, interview wave and country of residence, while FE regressions control for age, age squared and interview wave. 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.*

Table 5 First Stage Regression of the Binary Retirement Effect

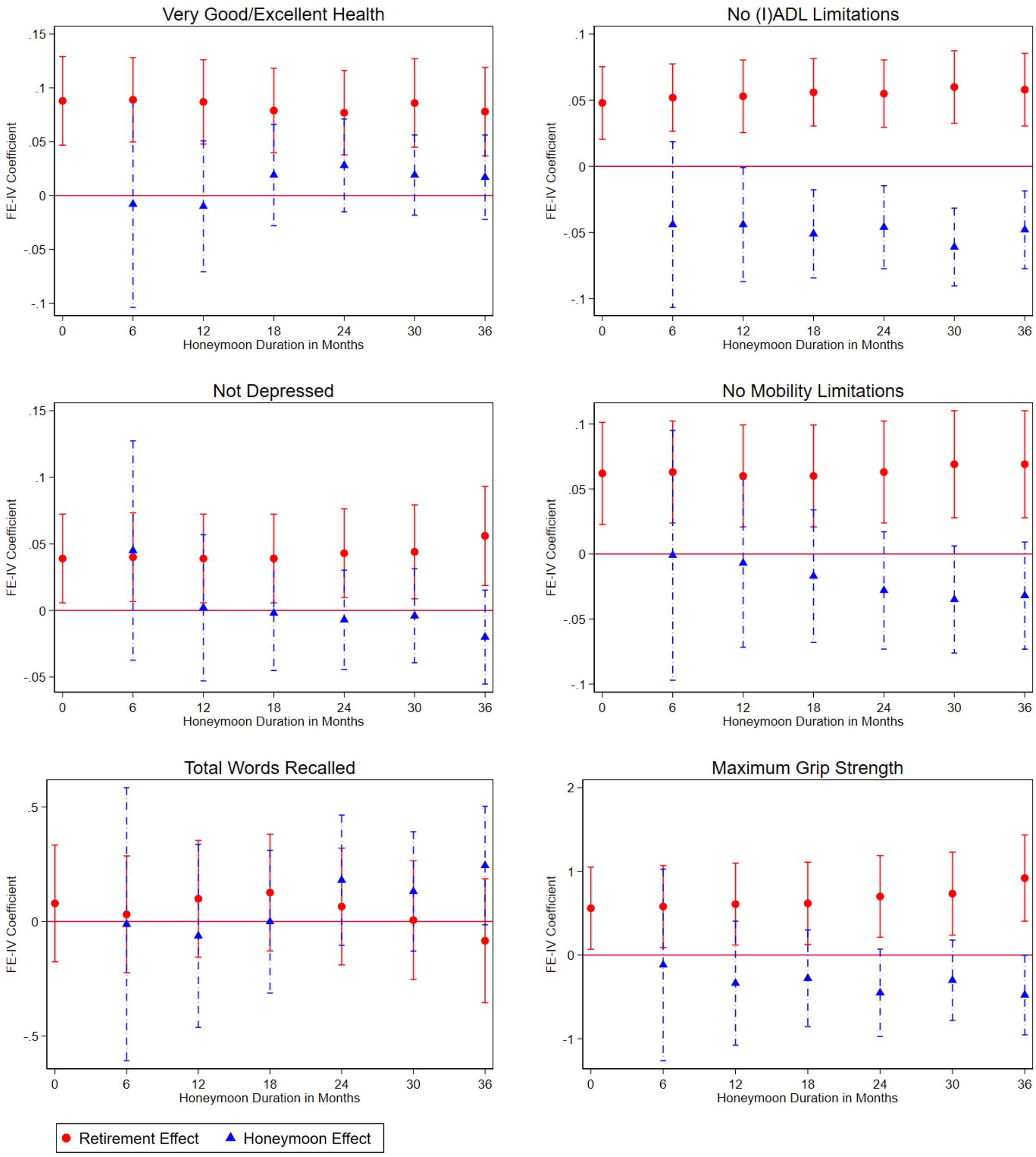
	(1)	(2)	(3)
Retired			
*over NRA*female*	0.390***		0.376***
	(0.011)		(0.011)
*over NRA*male*	0.290***		0.270***
	(0.010)		(0.010)
*over ERA*female*		0.253***	0.202***
		(0.011)	(0.011)
*over ERA*male*		0.240***	0.222***
		(0.010)	(0.010)
N	59 813	59 813	59 813

*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 interview wave and individual fixed effects. This first stage is for the analyses on Very Good / Excellent Health. The first stages in regressions with other dependent variables are extremely similar, with slightly different sample sizes. Significance levels: ***p<0.01, **p<0.05, *p<0.1.*

Columns (3) and (4) in Table 4 suggest that the effects of retiring at the NRA and ERA are in the same direction and of comparable magnitude. Some differences remain, which should be considered if only one of the age thresholds is used for the analysis. For example, a boost in subjective health appears more strongly when retiring at the ERA, while health improvements in terms of (I)ADL limitations and grip strength are concentrated among those retiring at the NRA. For the remainder of the analyses, only the results using both instruments jointly are presented. Retiring leads to an 8.8% increase in reporting very good/excellent health, which is quite large when compared to only 35% of individuals reporting very good or excellent health overall. Depression occurrence decreases by 3.9%, with 19.21% overall considered depressed. The probability to experience limitations in (I)ADL or mobility decrease by 5.0% and 6.2% respectively, with 12.8% and 39.3% of observations reporting limitations in (I)ADL and mobility respectively overall. Retirees are also significantly stronger, increasing their grip strength by 0.6 kg on average. Compared to the average grip strength (36.6 kg) and its standard deviation (11.7 kg), this effect appears less prominent than the other health benefits. An alternate way of looking at grip strength is to use an indicator variable for being above the thresholds used for the diagnosis of sarcopenia, as discussed in Bertoni et al. (2018): 20kg for women and 30kg for men. Using such an indicator shows that retirement significantly increases the likelihood to be above the threshold by 5.1%, supporting that the small increase in grip strength at retirement brings individuals above the threshold.

Having established a health improving binary effect of retirement on health, Figure 2 investigates whether these health improvements are mainly present in the honeymoon phase. Contrary to the previously discussed beliefs, no temporary health benefits are felt within the first 6-36 months of retirement. In many cases, the sign of the honeymoon phase effect is even negative, pointing towards more health issues in the short term – although these effects are not significant except for (I)ADL limitations (Table A-1).[4] The effect of retirement, even when controlling for the honeymoon phase, is stable, showing that there is a persistent health boost when retiring.

Figure 2: The Effect of Honeymoon and Retirement on Health



Note: These are the results of FE-IV regressions, using ERA and NRA jointly to instrument for the retirement decision. The 95% confidence intervals are shown. Positive coefficients (above the 0 line) imply a health improvement. Robust standard errors, clustered at the individual level, are used. All regressions control for age, age squared and interview wave.

---

[4] As the Kleibergen-Paap rk Wald F statistics in Table A-1 show, the instruments in all analyses can be considered strong.

5.2. Robustness Checks

Table 6 shows a set of robustness checks that each use a honeymoon phase of two years; results are similar when using other lengths. As the Kleibergen-Paap rk Wald F statistics in the table show, the instruments in all analyses are strong. Appendix Table A-2 shows the first stage results for the main analysis that incorporates the honeymoon phase.

First, we use alternative ways of clustering the standard errors. In the main analyses we clustered the standard errors are the individual level, which allows for arbitrary correlations within individuals. Columns (2), (3) and (4) show that results remain unchanged when we cluster at the household level, country*birth year*gender level, or country*gender level.[5]

Next, we analyze the robustness of our results against using alternate definitions of retirement, the instruments as well as alternate age trends. The regressions in column (5) use a purer definition of retirement, in which all individuals reporting to be retired yet still doing paid work are considered working. Most mechanisms thought to affect health due to retiring are based on lifestyle changes. Since individuals who continue working will experience fewer changes, they should also experience fewer health improvements. As predicted, the magnitude of the retirement effect increases in all six health measures. There is now evidence for health improvement during the honeymoon phase in terms of self-assessed health.

Results are also robust to redefining the instruments. In column (6), the instrument does not differ by gender, while the model in column (7) 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. Results are also robust to using different age specifications (columns (8) through (11)).

Another set of robustness checks includes various alternate sample selections. First, 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 to including those who ever report being unemployed, disabled, homemakers or do not declare their employment status as well as to including those who were self-employed or civil servants. Using different age spans (50-70 or 50-75) also yields similar results (Table A-3).

---

[5] Note that in the latter analysis the number of clusters is reduced to only 20. The small changes in coefficients when clustering at the household level are due to some individuals dropping out because of a missing household ID.

Table 6: Robustness Checks for Main Specification

		Alternative clustering				Instrument Definition		Different Age Specifications			
	Main	Household	Birth year* Country*Gender	Country*Gender	Alt. Ret. Definition	One IV	Country* Gender	Country* Age	Linear	Cubic	Quartic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Very good to excellent SAH											
retired	0.077***	0.076***	0.077***	0.077***	0.123***	0.079***	0.057***	0.094***	0.073***	0.087***	0.071**
	(0.020)	(0.020)	(0.020)	(0.021)	(0.033)	(0.021)	(0.018)	(0.021)	(0.019)	(0.029)	(0.034)
honeymoon	0.028	0.029	0.028	0.028	0.052**	0.022	0.006	0.017	0.022	0.019	0.015
	(0.022)	(0.022)	(0.021)	(0.022)	(0.022)	(0.022)	(0.018)	(0.022)	(0.021)	(0.026)	(0.026)
KP F-stat	[213]	[207]	[73]	[23]	[157]	[410]	[35]	[209]	[250]	[193]	[158]
No clinical depression											
retired	0.043**	0.042**	0.043**	0.043***	0.068**	0.038**	0.041***	0.033*	0.071***	0.026	0.024
	(0.017)	(0.017)	(0.018)	(0.014)	(0.028)	(0.017)	(0.015)	(0.018)	(0.016)	(0.026)	(0.030)
honeymoon	-0.007	-0.007	-0.007	-0.007	0.007	-0.012	-0.002	-0.007	0.013	0.004	0.005
	(0.019)	(0.020)	(0.021)	(0.013)	(0.019)	(0.019)	(0.016)	(0.020)	(0.019)	(0.024)	(0.024)
KP F-stat	[211]	[205]	[73]	[24]	[158]	[404]	[35]	[206]	[246]	[191]	[156]
Total Words Recalled (Cognitive Ability)											
retired	0.065	0.071	0.065	0.065	0.124	0.066	-0.158	0.260*	0.551***	-0.103	-0.053
	(0.130)	(0.132)	(0.159)	(0.166)	(0.211)	(0.131)	(0.114)	(0.133)	(0.125)	(0.188)	(0.219)
honeymoon	0.180	0.168	0.180	0.180	0.201	0.179	-0.030	0.082	0.512***	0.294*	0.305*
	(0.145)	(0.147)	(0.178)	(0.141)	(0.142)	(0.146)	(0.119)	(0.146)	(0.144)	(0.177)	(0.176)
KP F-stat	[204]	[198]	[71]	[23]	[153]	[392]	[34]	[199]	[238]	[185]	[151]
No (I)ADL Limitations											
retired	0.055***	0.055***	0.055***	0.055***	0.087***	0.057***	0.051***	0.051***	0.099***	0.014	0.014
	(0.014)	(0.014)	(0.014)	(0.011)	(0.022)	(0.014)	(0.012)	(0.014)	(0.014)	(0.021)	(0.024)
honeymoon	-0.046***	-0.047***	-0.046***	-0.046***	-0.026*	-0.046***	-0.043***	-0.043***	-0.016	-0.016	-0.017
	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)	(0.016)	(0.013)	(0.016)	(0.015)	(0.019)	(0.019)

KP F-stat	[213]	[207]	[73]	[23]	[158]	[408]	[35]	[208]	[249]	[192]	[157]
No Mobility Limitations											
retired	0.063***	0.063***	0.063***	0.063**	0.097***	0.061***	0.037**	0.056***	0.080***	0.049	0.053
	(0.020)	(0.021)	(0.020)	(0.030)	(0.033)	(0.020)	(0.018)	(0.021)	(0.019)	(0.030)	(0.035)
honeymoon	-0.028	-0.028	-0.028	-0.028	-0.008	-0.029	-0.028	-0.022	-0.017	-0.019	-0.018
	(0.023)	(0.023)	(0.023)	(0.025)	(0.023)	(0.023)	(0.019)	(0.023)	(0.022)	(0.028)	(0.028)
KP F-stat	[213]	[207]	[73]	[23]	[158]	[409]	[35]	[208]	[249]	[193]	[158]
Maximum Grip Strength											
retired	0.700***	0.707***	0.700**	0.700**	1.314***	0.433*	0.486**	0.563**	1.352***	0.636*	0.848**
	(0.249)	(0.250)	(0.348)	(0.350)	(0.406)	(0.258)	(0.213)	(0.256)	(0.235)	(0.360)	(0.423)
honeymoon	-0.451*	-0.443	-0.451	-0.451	-0.183	-0.513*	-0.437**	-0.427	0.043	-0.403	-0.351
	(0.266)	(0.270)	(0.352)	(0.362)	(0.263)	(0.278)	(0.221)	(0.270)	(0.263)	(0.321)	(0.320)
KP F-stat	[199]	[194]	[74]	[27]	[147]	[381]	[32]	[195]	[233]	[181]	[148]

*Note: N between 54 720 and 59 589 depending on outcome variable. These are the results of FE-IV regressions, where ERA and NRA are used jointly to instrument for the retirement decision. Honeymoon phase is considered to be two years. Positive coefficients imply a health improvement. Robust standard errors, clustered at the individual level unless otherwise indicated, are given in parentheses. Kleibergen-Paap (KP) rk Wald F-statistics are reported in brackets. All regressions control for age, age squared and interview wave. Significance levels: ***p<0.01, **p<0.05, *p<0.1.*

Another worry could be attrition due to death, which may lead to bias if it is related to retirement probability. The SHARE end of life questionnaire is used to identify the 959 individuals who have passed away (around 4.5% of the individuals included in the main analysis). In the first wave after they passed away, we assign these individuals 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. Three different retirement definitions are used, as changes to retirement between the last wave and death are not observed. First, the status reported in the last interview is used. 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. Results are robust (Table A-3).

In any longitudinal panel, attrition can also occur through a variety of other reasons beside death. If attrition occurs in a nonrandom way, this may bias results. Our strategy of individual fixed effects is a partial solution against this, as these control for time-invariant characteristics of individuals that are correlated with the probability of dropping out, while simultaneously affecting health outcomes. Nevertheless, a potential source of bias may remain if the propensity to dropping out is correlated with the size of the effect of retirement on health. To deal with this latter source of attrition, we conduct an additional analysis which shares similarities with that of Heller-Sahlgren (2017). We first run a probit of a dummy variable indicating whether respondents remained in the SHARE sample until the final wave that we included on age, age squared, female, number of children, years of education, whether the person had been a blue or white collar worker, country of residence, and our health outcomes – all as observed during the first wave in which a respondent was observed. We run this probit separately by wave in which the respondent first entered the SHARE survey, as the probability of attrition obviously greatly varies according to this variable. These regressions give us the probability of dropping out for each individual. We next assign a weight to each individual that corresponds to the ratio of the individual's probability of dropping out divided by the average probability for those respondents who enter SHARE for the first time in the same wave. (Hence, the average weight equals 1 for all respondents entering in the same wave.) We finally run weighted versions of our main regressions. Once more, the results are robust (Table A-3).

5.3. Heterogeneity

After having established the robustness of the main specification, the effect of retirement is analyzed within different populations (Table 7). 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 5.8%; it does not decrease for men. Maximum grip strength is significantly impacted by retirement only among women. While women gain 1.1 kg in strength in retirement, they lose some strength in the honeymoon phase. For men there is no significant effect, but the sign of the retirement coefficients goes into the opposite direction.

Columns (3) and (4) show the results when reducing the sample to individuals who reported being white-collar and blue-collar workers respectively. Other than expected, results suggest that blue- and white-collar experience similar health effects upon retirement. Blue-collar workers tend to have more health problems during the honeymoon phase, but experience similar if not greater health improvements afterwards. These results suggest that blue-collar workers go through a rough adjustment phase after retirement, before returning to better health.

Estimating the effect for further groups in terms of job characteristics leads to less precise results due to smaller sample sizes (Table 7). 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 retirement effect 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 (columns (5) and (6)). This is partially in contradiction to the results of Mazzona and Peracchi (2017) who find that retirement leads to an immediate health improvement only for people with physically demanding jobs, and then a gradual health decline for everyone. Like them, we find certain positive effects of retirement on cognitive performance for those with more physically demanding jobs, but unlike them we find no clear pattern for other outcomes. The differences in results may either come from a difference in model assumptions, or from a difference in definitions. Their model assumes a one-off jump at retirement and then a linear change over time till age 72 (their maximum age of observation), whereas our model assumes a two-year honeymoon phase and then a constant effect. And Mazzona and Peracchi classify people into higher vs lower physical burdens based on working conditions by ISCO-88 job classification whereas we rely on self-reports. Each of these capture slightly different dimensions of physical demandingness.

Our results also do not support the hypothesis 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 (columns (7) - (10)). Although there are slight differences depending on which health aspect is considered, no general trend can be identified.

Table 7: The Heterogeneity of the Effect of Retirement on Health

	Gender		Occupation		Physical Job		Time Pressure in Job		Freedom in Job	
	Female	Male	WC	BC	Yes	No	Yes	No	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Very Good / Excellent Health										
retired	0.071***	0.083***	0.076***	0.105**	0.128***	0.081**	0.150***	0.063*	0.147***	0.086***
	(0.026)	(0.032)	(0.023)	(0.046)	(0.042)	(0.035)	(0.043)	(0.036)	(0.056)	(0.031)
hnym.	0.043	0.007	0.043	-0.010	0.012	-0.006	-0.058	0.052	-0.065	0.026
	(0.028)	(0.033)	(0.026)	(0.039)	(0.053)	(0.048)	(0.054)	(0.049)	(0.078)	(0.040)
F-stat	243	181	154	53	46	60	51	52	23	82
N	27 781	32 032	39 366	18 832	12 429	17 056	14 681	14 801	7 383	22 087
Not Depressed										
retired	0.058**	0.021	0.061***	-0.017	0.040	0.048*	0.032	0.039	0.020	0.052**
	(0.025)	(0.023)	(0.019)	(0.041)	(0.036)	(0.027)	(0.036)	(0.028)	(0.049)	(0.024)
hnym.	-0.006	-0.016	0.021	-0.101**	0.018	0.010	0.045	-0.003	0.037	0.010
	(0.028)	(0.026)	(0.023)	(0.040)	(0.046)	(0.036)	(0.044)	(0.037)	(0.070)	(0.030)
F-stat	241	177	152	52	46	59	50	52	22	82
N	27 336	31 213	38 778	18 226	12 210	16 879	14 474	14 612	7 254	21 824
Total Words Recalled										
retired	0.120	0.024	0.023	0.220	0.574**	0.196	0.507*	0.229	0.377	0.334*
	(0.170)	(0.203)	(0.145)	(0.313)	(0.273)	(0.215)	(0.266)	(0.229)	(0.371)	(0.190)
hnym.	0.049	0.340	0.029	0.515*	0.040	-0.251	0.015	-0.293	-0.106	-0.104
	(0.194)	(0.219)	(0.176)	(0.274)	(0.343)	(0.291)	(0.346)	(0.291)	(0.514)	(0.244)
F-stat	236	171	147	51	45	57	50	51	22	79
N	26 922	30 645	37 897	18 138	11 955	16 514	14 226	14 240	7 135	21 323
No (I)ADL Limitations										
retired	0.058***	0.052***	0.049***	0.066*	0.025	0.037*	0.034	0.028	0.043	0.030
	(0.018)	(0.020)	(0.014)	(0.036)	(0.030)	(0.021)	(0.029)	(0.022)	(0.038)	(0.020)
hnym.	-0.041*	-0.053**	-0.036**	-0.077**	0.006	-0.037	0.005	-0.041	-0.015	-0.020
	(0.022)	(0.023)	(0.018)	(0.034)	(0.035)	(0.029)	(0.034)	(0.029)	(0.050)	(0.024)
F-stat	243	180	153	53	46	60	51	52	23	82
N	27 779	32 036	39 370	18 830	12 424	17 058	14 679	14 800	7 385	22 082
No Mobility Limitations										
retired	0.061**	0.063**	0.053**	0.132***	0.027	0.034	0.015	0.032	0.111**	0.007
	(0.027)	(0.031)	(0.023)	(0.048)	(0.041)	(0.036)	(0.044)	(0.035)	(0.056)	(0.031)
hnym.	-0.024	-0.034	-0.032	-0.021	0.046	-0.064	0.064	-0.081*	-0.121	0.013
	(0.032)	(0.033)	(0.027)	(0.046)	(0.052)	(0.047)	(0.053)	(0.047)	(0.078)	(0.039)
F-stat	244	180	154	53	46	60	51	52	23	82
N	27 783	32 034	39 370	18 832	12 426	17 057	14 682	14 798	7 385	22 083
Maximum Grip Strength										
retired	1.141***	-0.117	0.509*	1.147*	0.923*	0.038	-0.137	0.745*	0.568	0.382
	(0.267)	(0.450)	(0.263)	(0.659)	(0.473)	(0.394)	(0.464)	(0.410)	(0.621)	(0.345)
hnym.	-0.706**	-0.316	-0.419	-1.097*	-0.094	-0.502	0.275	-0.686	-0.120	-0.442
	(0.297)	(0.469)	(0.307)	(0.574)	(0.611)	(0.550)	(0.627)	(0.545)	(0.882)	(0.459)
F-stat	230	165	144	47	44	56	49	49	22	77
N	25 394	29 504	36 425	17 025	11 541	16 120	13 787	13 871	6 851	20 803

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

6. Discussion and Conclusion

Using SHARE data from waves 1, 2, 4, 5, and 6, we use a fixed effects instrumental variable approach to determine the causal effect of retirement 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 suggests a negative association between retirement and health. A large part of this negative association is driven by unobserved individual heterogeneity. 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.

Retiring either at the NRA or ERA significantly improves health. However, there are slight differences 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 decreases more strongly at the ERA, while improvements in (I)ADL limitations and maximum grip strength are only measured for retirement at the NRA. This may indicate that the Local Average Treatment Effect (LATE) differs between ages, as well as more generally between people who are induced to retire due to reaching the normal, vs the early retirement age threshold. Apparently, only retiring at a relatively older age due to reaching the normal retirement age leads to a greater grip strength and improvements in (I)ADL limitations.

On average, a health preserving effect of retirement is experienced, while there is no evidence for a temporary health-boosting honeymoon effect. Our results point towards the opposite occurring - retirees suffer from more mobility limitations and become weaker during the honeymoon phase. 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. Heterogeneity analyses show that both men and women experience health improvements upon retirement. Retirement also has a health preserving effect for blue- and white-collar workers. However, health worsening during the honeymoon phase occurs more among blue-collar workers, suggesting some difference in how white- and blue-collar workers transition into retirement. Other than hypothesized, it is not the case that those who worked in physically and psychologically straining jobs benefit more from retirement.

One main drawback of our approach is that the identified effect is a local average treatment effect (LATE) reflecting the effects for persons who are induced to retire due to reaching a retirement age threshold. No conclusions can be drawn about the impact retirement has on individuals who retire due to different reasons at another age. Nevertheless, we 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 lives, as those who retire due to other reasons will likely continue doing so in the future. 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. Mechanisms leading to worse health during the honeymoon phase should also be investigated. If a policy maker wants to target inequality among the elderly, it may also be important to consider that all workers seem to have a similar average 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, Jens; Schupp, Jürgen; Wagner, Gert G. (2011): Comparing the predictive power of subjective and objective health indicators: changes in hand grip strength and overall satisfaction with life as predictors of mortality. SOEPpapers on Multidisciplinary Panel Data Research, No. 398. DIW Berlin. DIW Berlin.

Atalay, K., Barrett, G. F., & Staneva, A. (2019). The effect of retirement on elderly cognitive functioning. *Journal of Health Economics*, *66*, 37-53.

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

Baum, Christopher F.; Schaffer, Mark E.; Stillman, Steven (2007): Enhanced routines for instrumental/generalized method of moments estimation and testing. In *The Stata Journal* 7 (4), pp. 465–506.

Behncke, Stefanie (2012): Does retirement trigger ill health? In *Health Economics* 21 (3), pp. 282–300. DOI: 10.1002/hec.1712.

Bertoni, M., Brunello, G., & Mazzarella, G. (2018). Does postponing minimum retirement age improve healthy behaviors before retirement? Evidence from middle-aged Italian workers. *Journal of Health Economics*, *58*, 215-227.

Bertoni, Marco; Maggi, Stefania; Weber, Guglielmo (2018): Work, retirement, and muscle strength loss in old age. In *Health Economics* 27 (1), pp. 115–128. DOI: 10.1002/hec.3517.

Bloemen, H., Hochguertel, S., & Zweerink, J. (2017). The causal effect of retirement on mortality: Evidence from targeted incentives to retire early. *Health Economics*, *26*(12), e204-e218.

Bonsang, Eric; Adam, Stéphane; Perelman, Sergio (2012): Does retirement affect cognitive functioning? In *Journal of Health Economics* 31 (3), pp. 490–501. DOI: 10.1016/j.jhealeco.2012.03.005.

Bonsang, E., & Klein, T. J. (2012). Retirement and subjective well-being. *Journal of Economic Behavior & Organization*, *83*(3), 311-329.

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, S. Zuber (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology.*

Bound, John; Waidmann, Timothy (2007): Estimating the Health Effects of Retirement. UM Working Paper 168. University of Michigan Retirement Research Center. University of Michigan Retirement Research Center.

Card, David; Dobkin, Carlos; Maestas, Nicole (2008): The impact of nearly universal insurance coverage on health care utilization: evidence from Medicare. In *The American Economic Review* 98 (5), pp. 2242–2258.

Celidoni, M., Dal Bianco, C., & Weber, G. (2017). Retirement and cognitive decline. A longitudinal analysis using SHARE data. *Journal of Health Economics*, *56*, 113-125.

Coe, Norma B.; Gaudecker, Hans-Martin von; Lindeboom, Maarten; Maurer, Jürgen (2012): The effect of retirement on cognitive functioning. In *Health Economics* 21 (8), pp. 913–927. DOI: 10.1002/hec.1771.

Coe, Norma B.; Lindeboom, Maarten (2008): Does retirement kill you? Evidence from early retirement windows. IZA Discussion Papers, No. 3817. Institute of Labor Economics. IZA - Institute of Labor Economics.

Coe, Norma B.; Zamarro, Gema (2011): Retirement effects on health in Europe. In *Journal of Health Economics* 30 (1), pp. 77–86. DOI: 10.1016/j.jhealeco.2010.11.002.

Dave, Dhaval; Rashad, Inas; Spasojevic, Jasmina (2008): The effects of retirement on physical and mental health outcomes. In *Southern Economic Journal* 75 (2), pp. 497–523.

Dwyer, Debra Sabatini; Mitchell, Olivia S. (1999): Health problems as determinants of retirement: Are self-rated measures endogenous? In *Journal of Health Economics* 18 (2), pp. 173–193.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of Health Economics*, *43*, 1-12.

Fé, E., & Hollingsworth, B. (2016). Short-and long-run estimates of the local effects of retirement on health. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 1051-1067.

Godard, M. (2016). Gaining weight through retirement? Results from the SHARE survey. *Journal of Health Economics*, *45*, 27-46.

Gorry, Aspen; Gorry, Devon; Slavov, Sita Nataraj (2018): Does retirement improve health and life satisfaction? In *Health Economics* 27 (12), pp. 2067–2086. DOI: 10.1002/hec.3821.

Grip, Andries De; Lindeboom, Maarten; Montizaan, Raymond (2012): Shattered dreams: the effects of changing the pension system late in the game. In *The Economic Journal* 122 (559), pp. 1–25. DOI: 10.1111/j.1468-0297.2011.02486.x.

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

Grøtting, M. W., & Lillebø, O. S. (2020). Health effects of retirement: evidence from survey and register data. *Journal of Population Economics*, *33*(2), 671-704.

Hagen, J. (2018). The effects of increasing the normal retirement age on health care utilization and mortality. *Journal of Population Economics*, *31*(1), 193-234.

Hallberg, Daniel; Johansson, Per; Josephson, Malin (2015): Is an early retirement offer good for your health? Quasi-experimental evidence from the army. In *Journal of Health Economics* 44, pp. 274–285. DOI: 10.1016/j.jhealeco.2015.09.006.

Hank, K.; Jürges, H.; Schupp, J.; Wagner, G. G. (2009): Isometrische Greifkraft und sozialgerontologische Forschung. Ergebnisse und Analysepotentiale des SHARE und SOEP. In *Zeitschrift fur Gerontologie und Geriatrie* 42 (2), pp. 117–126. DOI: 10.1007/s00391-008-0537-8.

Heller-Sahlgren, Gabriel (2017): Retirement blues. In *Journal of Health Economics* 54, pp. 66–78. DOI: 10.1016/j.jhealeco.2017.03.007.

Herbay, Axel von (2014): Otto von Bismarck is not the origin of old age at 65. In *The Gerontologist* 54 (1), p. 5. DOI: 10.1093/geront/gnt111.

Hernaes, E., Markussen, S., Piggott, J., & Vestad, O. L. (2013). Does retirement age impact mortality?. *Journal of Health Economics*, *32*(3), 586-598.

Hofäcker, Dirk (2015): In line or at odds with active ageing policies? Exploring patterns of retirement preferences in Europe. In *Ageing and Society* 35 (7), pp. 1529–1556. DOI: 10.1017/S0144686X1400035X.

Idler, Ellen L.; Benyamini, Yael (1998): Self-rated health and mortality: a review of twenty-seven community studies. In *Journal of Health and Social Behavior* 38 (1), pp. 21–37.

Insler, Michael (2014): The health consequences of retirement. In *The Journal of Human Resources* 49 (1), pp. 195–233. DOI: 10.1353/jhr.2014.0000.

Kesavayuth, D., Rosenman, R. E., & Zikos, V. (2020). The dynamic effects of retirement on well-being. *Applied Economics*, *52*(54), 5921-5935.

Kim, Jungmeen E.; Moen, Phyllis (2002): Retirement transitions, gender, and psychological well-being: a life course, ecological model. In *Journal of Gerontology Series B: Psychological Sciences & Social Sciences* 57 (3), pp. 212–222.

Leong, Darryl P.; Teo, Koon K.; Rangarajan, Sumathy; Lopez-Jaramillo, Patricio; Avezum, Alvaro; Orlandini, Andres et al. (2015): Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology (PURE) study. In *The Lancet* 386 (9990), pp. 266–273. DOI: 10.1016/S0140-6736(14)62000-6.

Lucifora, C., & Vigani, D. (2018). Health care utilization at retirement: The role of the opportunity cost of time. *Health Economics*, *27*(12), 2030-2050.

Mazzonna, Fabrizio; Peracchi, Franco (2012): Ageing, cognitive abilities and retirement. In *European Economic Review* 56 (4), pp. 691–710. DOI: 10.1016/j.euroecorev.2012.03.004.

Mazzonna, Fabrizio; Peracchi, Franco (2017): Unhealthy retirement? In *The Journal of Human Resources* 52, pp. 128–151.

Neuman, Kevin (2008): Quit your job and get healthier? The effect of retirement on health. In *Journal of Labor Research* 29 (2), pp. 177–201. DOI: 10.1007/s12122-007-9036-8.

Nielsen, N. F. (2019). Sick of retirement?. *Journal of Health Economics*, *65*, 133-152.

Nishimura, Yoshinori; Oikawa, Masato; Motegi, Hiroyuki (2018): What explains the difference in the effect of retirement on health? Evidence from global aging data. In *Journal of Economic Surveys* 32 (3), pp. 792–847. DOI: 10.1111/joes.12215.

OECD (2015): Ageing and employment policies. Denmark 2015 - working better with age. In *OECD Publishing*.

Petrou, Stavros; Kupek, Emil (2008): Social capital and its relationship with measures of health status: evidence from the Health Survey for England 2003. In *Health Economics* 17 (1), pp. 127–143. DOI: 10.1002/hec.1242.

Prince, M. J.; Freischies, F.; Beekman, A. T. F.; Fuhrer, R.; Jonker, C.; Kivela, S.-L. et al. (1999): Development of the EURO-D scale - a European Union initiative to compare symptoms of depression in 14 European centres. In *British Journal of Psychiatry* 174 (4), pp. 330–338.

Rohwedder, Susann; Willis, Robert J. (2010): Mental retirement. In *Journal of Economic Perspectives* 24 (1), pp. 119–138. DOI: 10.1257/jep.24.1.119.

Shai, O. (2018). Is retirement good for men's health? Evidence using a change in the retirement age in Israel. *Journal of Health Economics*, *57*, 15-30.

Stock, James H.; Yogo, Motohiro (2005): Testing for weak instruments in linear IV regression. In Donald W. K. Andreas, James H. Stock (Eds.): Identification and Inference for Econometric Models. Essays in Honor of Thomas Rothenberg: Cambridge University Press, pp. 80–108.

Thomas, Patricia A. (2011): Gender, social engagement, and limitations in late life. In *Social Science & Medicine* 73 (9), pp. 1428–1435. DOI: 10.1016/j.socscimed.2011.07.035.

World Bank (2016b): Life expectancy at birth, total (years). World Development Indicators. Edited by The World Bank Group. The World Bank Group. Available online at https://data.worldbank.org/indicator/SP.DYN.LE00.IN.

Zantinge, Else M.; van den Berg, Matthijs; Smit, Henriëtte A.; Picavet, H. Susan J. (2014): Retirement and a healthy lifestyle: opportunity or pitfall? A narrative review of the literature. In *European Journal of Public Health* 24 (3), pp. 433–439. DOI: 10.1093/eurpub/ckt157.

Zhang, Y., Salm, M., & van Soest, A. (2018). The effect of retirement on healthcare utilization: Evidence from China. *Journal of Health Economics*, *62*, 165-177.

Appendix

Table A-1 The Effect of Honeymoon and Retirement on Health – Testing Different Honeymoon Phases

	Duration of Honeymoon Phase (in months)						
	0	**6**	**12**	**18**	**24**	**30**	**36**
Very Good / Excellent Health							
retired	0.087***	0.089***	0.087***	0.079***	0.077***	0.085***	0.078***
	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)
honeymoon		-0.010	-0.012	0.018	0.025	0.018	0.016
		(0.049)	(0.031)	(0.024)	(0.022)	(0.020)	(0.020)
KP F-stat	829	54	113	174	213	260	269
N	59 583	59 583	59 583	59 583	59 583	59 583	59 583
Not Depressed							
retired	0.038**	0.039**	0.038**	0.038**	0.042**	0.042**	0.055***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.019)
honeymoon		0.044	0.003	-0.001	-0.007	-0.004	-0.020
		(0.042)	(0.028)	(0.022)	(0.019)	(0.018)	(0.018)
KP F-stat	827	54	111	173	210	256	266
N	58 355	58 355	58 355	58 355	58 355	58 355	58 355
Total Words Recalled							
retired	0.076	0.087	0.094	0.123	0.061	0.002	-0.083
	(0.130)	(0.130)	(0.130)	(0.130)	(0.130)	(0.132)	(0.138)
honeymoon		-0.011	-0.056	-0.004	0.179	0.129	0.233*
		(0.303)	(0.204)	(0.160)	(0.145)	(0.133)	(0.132)
KP F-stat	810	53	107	168	203	249	259
N	57 371	57 371	57 371	57 371	57 371	57 371	57 371
No (I)ADL Limitations							
retired	0.048***	0.052***	0.053***	0.056***	0.055***	0.060***	0.058***
	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)
honeymoon		-0.045	-0.045**	-0.049***	-0.045***	-0.061***	-0.047***
		(0.032)	(0.022)	(0.017)	(0.016)	(0.015)	(0.015)
KP F-stat	830	54	112	174	212	259	269
N	59 585	59 585	59 585	59 585	59 585	59 585	59 585
No Mobility Limitations							
retired	0.061***	0.062***	0.059***	0.060***	0.062***	0.068***	0.069***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)
honeymoon		-0.003	-0.010	-0.018	-0.029	-0.036*	-0.033
		(0.050)	(0.033)	(0.026)	(0.023)	(0.021)	(0.021)
KP F-stat	829	54	112	174	212	259	269
N	59 589	59 589	59 589	59 589	59 589	59 589	59 589
Maximum Grip Strength							
retired	0.570**	0.587**	0.615**	0.622**	0.704***	0.739***	0.925***
	(0.251)	(0.250)	(0.250)	(0.251)	(0.249)	(0.254)	(0.263)
honeymoon		-0.085	-0.316	-0.249	-0.434	-0.284	-0.459*
		(0.584)	(0.378)	(0.295)	(0.267)	(0.246)	(0.243)
KP F-stat	795	52	106	163	199	243	252
N	54 720	54 720	54 720	54 720	54 720	54 720	54 720

*Note: These are the results of FE-IV regressions, where ERA and NRA are used jointly to instrument for the retirement decision. Positive coefficients imply a health improvement. 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 and interview wave. Significance levels: ***p<0.01, **p<0.05, *p<0.1.*

Table A-2 First Stage of an FE-IV Regression with a Honeymoon Phase of Two Years

Retired	
*24 or more months over NRA*female*	-0.178***
	(0.014)
*24 or more months over NRA*male*	-0.182***
	(0.012)
*24 or more months over ERA*female*	0.092***
	(0.014)
*24 or more months over ERA*male*	0.051***
	(0.014)
*Up to 24 months over NRA*female*	0.202***
	(0.017)
*Up to 24 months over NRA*male*	0.128***
	(0.014)
*Up to 24 months over ERA*female*	0.126***
	(0.013)
*Up to 24 months over ERA*male*	0.090***
	(0.013)
N	59,813

*Note: This first stage is for the analysis on Very Good / Excellent Health. The first stages in regressions with other dependent variables are extremely similar, with slightly different sample sizes. Significance levels: ***p<0.01, **p<0.05, *p<0.1.*

Table A-3 Further Robustness Checks

	w/o big changes	Ever unmpl. incl.	Unemployed incl.	incl. Self-employed and civil servants	50-70 year olds	50-75 year olds	Including Deceased: Ret. Stat. Last Wave	Including Deceased: Ret. If Age≥ERA	Including Deceased: Ret. If Age≥NRA	Weighted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Very Good / Excellent Health										
retired	0.093***	0.081***	0.078***	0.073***	0.072**	0.078***	0.077***	0.075***	0.072***	0.082***
	(0.022)	(0.020)	(0.018)	(0.018)	(0.032)	(0.021)	(0.021)	(0.020)	(0.020)	(0.022)
hnym.	0.006	0.025	0.031	0.024	0.020	0.021	0.027	0.028	0.028	0.026
	(0.023)	(0.021)	(0.020)	(0.019)	(0.032)	(0.024)	(0.021)	(0.021)	(0.021)	(0.023)
KP F-stat	[175]	[229]	[240]	[276]	[137]	[189]	[216}	[216]	[219]	[186]
N	48 622	61 760	65 259	91 051	41 713	52 176	61 298	61 298	61 298	55 218
Not Depressed										
retired	0.037**	0.038**	0.045***	0.043***	0.052*	0.037**	0.054***	0.052***	0.052***	0.038*
	(0.019)	(0.017)	(0.016)	(0.015)	(0.029)	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)
hnym.	-0.009	-0.008	-0.001	-0.018	-0.005	0.003	-0.010	-0.010	-0.010	-0.020
	(0.020)	(0.019)	(0.019)	(0.017)	(0.029)	(0.022)	(0.020)	(0.020)	(0.020)	(0.022)
KP F-stat	[173]	[225]	[237]	[271]	[135]	[186]	[214]	[214]	[217]	[185]
N	47 565	60 490	63 899	89 188	41 041	51 198	59 994	59 994	59 994	54 434
Total Words Recalled										
retired	-0.043	0.083	0.060	0.216*	0.200	-0.010	0.035	0.030	0.029	0.145
	(0.138)	(0.127)	(0.118)	(0.116)	(0.208)	(0.138)	(0.134)	(0.131)	(0.131)	(0.144)
hnym.	0.161	0.185	0.194	0.206	-0.011	0.268*	0.225	0.226	0.224	0.254
	(0.150)	(0.139)	(0.137)	(0.129)	(0.214)	(0.160)	(0.147)	(0.147)	(0.146)	(0.161)
KP F-stat	[173]	[218]	[229]	[264]	[131]	[181]	[208]	[207]	[210]	[177]
N	47 626	59 483	62 855	87 674	40 282	50 306	59 019	59 019	59 019	53 443
No (I)ADL Limitations										
retired	0.056***	0.061***	0.066***	0.047***	0.015	0.047***	0.060***	0.059***	0.059***	0.059***
	(0.015)	(0.013)	(0.013)	(0.012)	(0.023)	(0.015)	(0.014)	(0.014)	(0.014)	(0.016)
hnym.	-0.051***	-0.046***	-0.042***	-0.022	-0.007	-0.034*	-0.044***	-0.044***	-0.044***	-0.049***
	(0.017)	(0.015)	(0.015)	(0.014)	(0.022)	(0.017)	(0.017)	(0.017)	(0.016)	(0.018)
KP F-stat	[175]	[228]	[239]	[275]	[136]	[188]	[216]	[216]	[219]	[186]
N	48 630	61 762	65 261	91 058	41 705	52 172	61 299	61 299	61 299	55 214
No Mobility Limitations										
retired	0.056**	0.069***	0.069***	0.064***	0.051	0.064***	0.065***	0.064***	0.064***	0.080***
	(0.022)	(0.020)	(0.019)	(0.017)	(0.034)	(0.022)	(0.021)	(0.020)	(0.020)	(0.023)
hnym.	-0.028	-0.036*	-0.030	-0.038*	-0.009	-0.027	-0.024	-0.023	-0.023	-0.046*
	(0.024)	(0.022)	(0.022)	(0.020)	(0.035)	(0.026)	(0.023)	(0.023)	(0.023)	(0.025)
KP F-stat	[175]	[228]	[239]	[275]	[137]	[188]	[216]	[216]	[219]	[186]
N	48 633	61 766	65 265	91 064	41 714	52 180	61 301	61 301	61 301	55 217
Maximum Grip Strength										
retired	0.842***	0.784***	0.788***	0.637***	0.769*	0.660**	0.797***	0.756***	0.739***	0.823***
	(0.261)	(0.244)	(0.228)	(0.221)	(0.409)	(0.262)	(0.270)	(0.265)	(0.265)	(0.285)
hnym.	-0.580**	-0.429*	-0.470*	-0.381	-0.427	-0.375	-0.165	-0.154	-0.172	-0.459
	(0.273)	(0.255)	(0.253)	(0.234)	(0.396)	(0.292)	(0.290)	(0.290)	(0.289)	(0.299)
KP F-stat	[169]	[213]	[224]	[258]	[126]	[177]	[202]	[202]	[205]	[174]
N	45 502	56 732	59 976	83 714	38 730	48 248	56 252	56 252	56 252	51 874

*Note: These are the results of FE-IV regressions, where ERA and NRA are used jointly to instrument for the retirement decision. Honeymoon phase is considered to be two years. Positive coefficients imply a health improvement. 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 and interview wave. Significance levels: ***p<0.01, **p<0.05, *p<0.1.*