Journal Pre-proofs

The influence of old-age retirement on health: Causal evidence from the Finnish register data

T. Kuusi, P. Martikainen, T. Valkonen

Please cite this article as: T. Kuusi, P. Martikainen, T. Valkonen, The influence of old-age retirement on health: Causal evidence from the Finnish register data, *The Journal of the Economics of Ageing* (2020), doi: https://doi.org/10.1016/j.jeoa.2020.100257

The influence of old-age retirement on health: Causal evidence from the Finnish register data

Kuusi, T., Martikainen, P., Valkonen, T.[1]

Abstract

We quantify the impact of old-age retirement on health using longitudinal Finnish register data for the period 2000–2012. The data allows for a strict isolation of the effects of transition from work to retirement for both mental and physical health indicators. We use the lowest statutory eligibility age for full old-age pensions, 63 years, as an instrument in FE-IV estimation to ensure causal inference. We find that (1) retirement at age 63 moderately decreases the use of antidepressants, especially for women. The effect is sharp, while it is somewhat reversed in the later years; (2) the beneficial effects of retirement on the cardiovascular and musculoskeletal conditions are smaller and more diffused; (3) for occupational classes, our results show a reduction in antidepressant use for women in almost all occupations, whereas for men it is significant for manual workers and farmers; (4) we find stronger declines in the anti-depressant use among men and women who retire from low- and mid-income work, as well as more robust decrease in the risk of cardiovascular diseases for high-income and non-single; and (5) our test of external validity shows that the beneficial effects in antidepressant use can be extended to apply to most Finns retiring at ages 62–64.

Keywords: health, health behaviour, retirement, retirement policies, demographic economics
JEL-codes: I10, I12, J26, J18

1

1. Introduction

The current trend of declining old-age mortality rates has severely challenged the financial sustainability of the public pension schemes in almost all the high-income countries. A yet-unresolved question is how far the retirement ages can be increased without significant effects on the health of elderly workers and retirees; this is an issue of both individual welfare and public finances. While finding answers has been a goal of a large body of research literature, a great deal of uncertainty remains in the subject. The research designs are challenged by ambiguous direction of causality and heterogeneous, often unobserved characteristics of retirees. Furthermore, the empirical literature typically builds on local average treatment effects (LATE) that identify the treatment effects for only a minority of the retirees.

In this paper, we introduce new estimates for the health effects of retirement by using the individual Finnish population, retirement and health registers for the period 2000–2012. Our focus is on retirements after the 2005 pension reform while the full data window allows us to control for the trend effects of ageing prior and after the retirement. We measure the influence of retirement on mental health through purchases of antidepressants and the impact of retirement on physical health by hospital visits associated with cardiovascular or musculoskeletal diseases. We show that our results are not likely to be subject of the typical identification problems in the literature, and thus provide a good benchmark of the health impacts of retirement.

The detailed data allows us to carefully analyse the characteristics of retirees. Our health data comprise information on all prescription medications and contacts with the hospital care system of the studied individuals. Therefore, we avoid the problems related to using self-assessed health measures and non-response or loss to follow-up biases. Furthermore, we identify persons that have retired from full-time work to full retirement, with the aim of capturing all the key elements of the retirement transition (end of work, more leisure, less income). A caveat in the previous literature has been that the observed retirements often involve cases in which retirees have already ceased working due to unemployment or disability before retirement, or they may have continued working after starting to draw pensions. In these cases, results may be ambiguous due to large de facto heterogeneity in the key elements of retirement. Moreover, we observe the health outcomes for different income and socio-economic groups for which qualities of work and retirement from it may be very different.

We make use of the Finnish institutional set-up that is suitable for a quasi-experimental research design. In particular, we address the health–retirement endogeneity problem using the instrumental variable (IV) approach. The endogeneity problem arises because changes in health may also affect the decision to retire and this may easily result in biased estimates of the health effects of retirement. Following the recent literature, we use the lowest statutory eligibility age for full old-age pensions as an instrument for the actual retirement age. In the Finnish earnings-related first pillar old-age pension scheme, the statutory earliest full pension retirement age of 63 years provides particularly strong incentives for retirement from work to old-age pension, and indeed it is the most common route out of labor markets. While the statutory age is a good predictor of the true retirement age during the studied period, it is not likely by itself to affect health. Thus, the lowest eligibility age provides a strong and valid instrument.

The pension scheme at our focus dictates a large fraction of the retirement income and retirement decisions in Finland. Most of the individuals working in the private sector until old-age pension retired between the lower and the upper age limits of the flexible retirement window of 62–67 years in our data. There are simple rules for pension accrual and actuarial adjustment for postponing withdrawals. Moreover, the replacement rates and old-age poverty are on average OECD levels and thus, the Finnish system provides a framework for a representative case in terms of the income effects of retirement.

Based on our analysis, we provide new and interesting findings on the effects of retirement. We show that retirement decreases the use of antidepressants, whereas the effect on physical conditions is smaller and more diverse. For socio-economical classes, our results show a reduction in antidepressant use for women in almost all occupations, whereas for men the effect is significant only for manual workers and farmers, while the effects are much smaller for the other occupational categories. Moreover, we find stronger declines in the anti-depressant use among men and women who retire from low- and mid-income work, as well as more robust decrease in the risk of cardiovascular diseases for men who work in the high-income jobs or whose family status is other than single. For women, we also find solid evidence that the risk of using antidepressants falls sharply after retirement, while the effect is somewhat reversed in the later years.

Moreover, our detailed data set allows us also to address the generalisability of our results. A key challenge of current retirement health literature is that IV methodology only provides the LATE at the exact eligibility age. Typically, a jump in the probability of retirement as a result of reaching the statutory age limit (the first stage) suggest that only minority of the population will comply with the rule – our study is no exception. As Figure 1 shows, in the Finnish case the increase in the probability of retirement due to reaching age 63 is at best around 30%, albeit that it is still a relatively strong effect when compared to the previous literature.[2] As some people choose to retire before the statutory retirement age, and others who reach it forgo the possibility of retiring, the decision to retire provides a case of two-sided noncompliance. If there is evidence against the homogeneity of the expected (marginal) treatment effects in the population, the group of compliers is selected and the selection matters for the outcome, thus compromising the generalisability of the results.
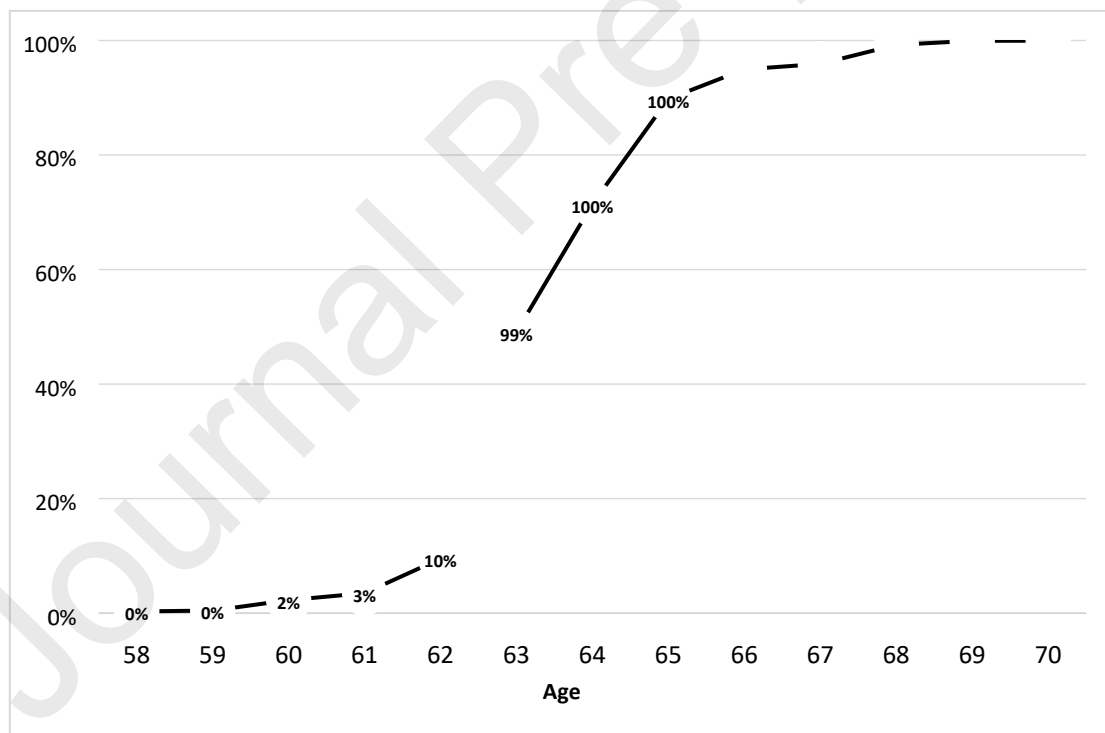
Figure 1. Cumulative distribution of retirement in the Finnish private sector 2005-2012. The figure shows the average % of retirees in the group of persons who retire from work to old-age pension by age. Note: More information concerning the definition of the group is given in Section 3.

---

[2] The figure shows the percentage of retirees in the group of persons that have a pure retirement transition from work to old-age pension (see Section 3 for definition), while the same also holds true for all old-age retirements.

We test the heterogeneity of the marginal effects of retirement on our health variables by applying the framework proposed by Brinch et al. (2017) in different segments of the population. We compare the health outcomes of those who retire at 63 with those who retire later or earlier, and build a rejection test statistic for treatment homogeneity.

Our general finding is that the homogeneity assumption can be rejected when comparing health outcomes far from the threshold age and if the time-invariant individual-level heterogeneity is not properly accounted for. Instead, in the age window of 62–64 years, and after considering the time-invariant effects, we do not find evidence against treatment homogeneity in the case of antidepressant use and hospital episodes due to musculoskeletal diseases. Thus, we can safely generalise our LATE results concerning these health outcome variables to the broader age window. This is important, as 68 % of retirements from work to old-age pensions occur at this age interval[3]. On the other hand, we find some evidence suggesting that the effect of retirement on the cardiovascular hospital treatments is heterogeneous in a manner that may result in upward bias in the effect of retirement. Moreover, there are some socioeconomic groups for which the homogeneity of the treatment effect is rejected.

This article is organized as follows. In Section 2, we take stock on the previous literature. Section 3 introduces our data, while Section 4 discusses our research methodology. In Section 5, we report our main results, and in Section 6 we address robustness of our results to different specifications and typical problems in statistical inference. Section 7 concludes.

2. Literature

The impacts of retirement on health have long been studied in the public health literature. Two summarising surveys by Maimaris et al. (2010) and van der Heide et al. (2013) report that retirement is good for mental health. Instead, the results of studies on cognitive skills reported by Meng et al. (2017), or on physical health, vary both with sign and intensity.[4]

Health economists joined the research tradition later and contributed by emphasising causal analysis using quasi-experimental designs and their related methods (regression discontinuity, differences-in-differences and instrumental variables; see Charles 2004 and Heller-Sahlgren 2017). A replication study (Nishimura et al. 2018) found that the choice of statistical methods and cofounders has a strong influence on the results.

The latter type of quasi-experimental studies shows that retirement effects are attenuated (Rohwedder and Willis 2010, Mazzonna and Perachhi 2012, Bonsang et al. 2012, Tumino 2016, Nishimura et al. 2017) or have no impact (Coe and Zamarro 2011, Coe et al. 2012, Kajitani et al. 2016) on cognitive skills. A replication study by Fonseca (2017) detects that controlling for country-specific differences strongly diminishes retirement effects.

Results regarding mental health show that the impact of retirement is either positive (Atalay and Barrett 2014, Zhu 2016, Kolodziej and Garcia-Gomez 2019, Oshio and Kan 2017, Belloni et al. 2016 [for men] and Mazzonna and Perachhi 2017 [for physically demanding jobs]) or is not significant (Behncke 2012, Coe and Zamarro 2011, Mokyr Horner and Cullen 2016). Heller-Sahlgren (2017) does not find immediate effects, but the long-term consequences are strongly negative.

---

[3] It is noteworthy that the number is very similar to the share (67%) of all old-age pension retirements at this age window in the corresponding time period.

[4] Grossman's (1972) human capital theory does not provide much help here. Investments in health do not increase wage income after retirement (only the consumption value of health remains), which reduces incentives to use time or money to this end. In contrast, retirement abolishes the cost of the investment in terms of lost wage income; the net effect is not known.

The main result from the previous studies analysing the physical health effects of retirement is that health improves. Insler (2014) shows that health gets better in the USA. Atalay and Barrett (2014) and Zhu (2016) study Australian women and detect positive results on a wide selection of health indicators. Coe and Zamarro (2011) reach similar results with the European Survey of Health, Ageing and Retirement in Europe (SHARE) data.

Some recent studies have discovered different responses for different groups. Mazzonna and Peracchi (2017) find that health improves in physically demanding occupations, but in occupations where physical stress is low, retirement may damage health. Hessel (2016) reports that retirement is beneficial to physical health in all the other groups, but not for lower-educated women. Eibich (2015) detects that lower-educated individuals benefit from better physical health, but the estimate for mental health is not significant. Highly educated individuals experience the opposite results. Negative effects have also been detected. Behncke (2012) shows that retirement increases chronic conditions and weakens self-assessed health in the UK.

Most of the studies mentioned above use either self-assessed health or health indices calculated using survey data as response variables. Several studies that assess the effects of retirement on registered mortality have found no effects (Hernaes et al. 2013), Nielsen (2018) and Hagen (2018), but Fitzpatrick and Moore (2017) claim that retirement increases male mortality in the USA. On the other hand, Bloemen et al. (2017) and Hallberg et at. (2015) detected that early retirement lowers mortality.

There is also a branch of literature that analyse the effects of retirement on health care use. The issue is relevant, fox example, when the effects of higher retirement ages on public finances are studied. A change in the use can stem from several reasons, such as a change in the access to care, lower opportunity costs of time used in health care investments after retirement, or the influence of retirement on health. The health effects of retirement are more likely to explain the change in health care use if the cause of the health care contact is severe illness or, correspondingly, if the measure of use is overnight stays in hospital instead of contacts to general practitioner. A change in access to primary care after retirement may emerge because of lost eligibility to employees' health insurance, or, as in Finland, because of lost access to occupational health care.

Zhang et al. (2018) finds that health care use increases after retirement markedly in China, which is likely to be caused mainly by deteriorated health, but there is also indirect evidence on the influence of lower opportunity costs. Lucifora and Vigani (2018) observed that retirement increases the probability of visiting a doctor and the number of doctor's visits. Bíró (2016) finds that the visits increase even conditional on health. On the other hand, Nielsen (2018) finds no effect of statutory or early retirement age on physical health measured with comorbidity or mortality in Danish register data but the use of the popular early retirement route seems to reduce general practitioner (GP) visits and hospitalisation in Denmark. Gorry, et al. (2018) found that health care use measured with several indicators decline after retirement in the U.S.A. Thus, the empirical evidence on the influence of lower opportunity costs is unclear and mixed with true changes in health.

Three recent studies focus on the impact of retirement on mental health using Finnish data. Oksanen et al. (2011) observe that employees who retire at the statutory age and those who retire early because of mental health problems reduce their purchases of antidepressants after retirement. Laaksonen et al. (2012) report a similar outcome for the use of antidepressants but an increased use of hypnotics and sedatives in the case of mental health-related retirement. There is no impact on medication use among old-age retirees. Leinonen et al. (2013) observe no impact on the use of antidepressants around old-age retirement, but they find a decline in the use of antidepressants for those who retire early because of mental health reasons. These studies do not use quasi-experimental designs and related methods.

The variation in the previous results listed above can arise from many sources. A comparison of recent studies is provided in Table A1 in the Appendix. These studies use dominantly either FE-IV models or regression discontinuity approach. Introducing individual fixed effects removes the time-invariant unobservable characteristics that may create a link between health and retirement. Correspondingly, use of IV-models or RDD analysis considers reverse causality. However, use of these methods often constraint the studied group to cover a small subpopulation, which limits the external validity of the results.

In this article, we address several potential caveats in identifying causal effects that may result in differences between our and the earlier results. We use standard FE-IV methodology, which capitalize on the earliest full pension retirement age as instrument to true retirement. Our methodological contribution is to expand the external validity to the result to the ages when most of the Finns retire. We also run a large amount of robustness tests to show, for example that reaching age 63 does not indicate, per se, marked changes in the health outcomes, as shown by the analysis of persons that do not retire at that age.

Our study is based on representative yearly individual panel data on objective health indicators, demographics, labour force status, socio-economic status, income and retirement scheme compiled from Finnish registers. The quality of data is high compared to most of the studies.

Measures of health vary in previous studies from self-assessed health and self-reported diseases to register-based detailed diagnoses and from use of health services to mortality. Data may be collected biannually, as in SHARE, yearly, or in case of registers, even monthly. Our study uses yearly purchases of prescription medicines and over-night stays in hospital.

Measured retirement may be self-reported or register-based, take place early, or at a statutory age, and the working intensity before retirement varies. In many countries, it is also common to work while drawing a pension. As opposed to many of the previous papers, we can isolate the pure retirement transition from work to old-age pension. This group is the relevant one when we consider the health effects of a higher retirement age. For those who are not working until the statutory retirement age, transition to non-employment has taken place earlier, and the income they receive may increase or decrease when eligibility for old age pension begins

3. Data
3.1. Study population
This study is based on a linked register-based 11% random sample of the population residing in Finland at the end of any of the years between 2000–2012. We study the effects of retirement on health after the 2005 pension reform while the data window allows us to control for the trend effects of ageing prior and after the retirement. The data originate from various registers covering the whole Finnish population. Sociodemographic data come from the Labour Market Data File maintained by Statistics Finland, data on hospital visits from the National Hospital Discharge Register held by the Finnish National Institute for Health and Welfare and information on purchases of prescription medication from retail pharmacies from the Finnish National Prescription Register held by the National Social Insurance Institution. We follow our study participants annually as many of the measurements – such as income – are only available on an annual basis.

Statistics Finland use personal identification numbers to combine data from different registers; the register holders have approved the use of the data for scientific research. Data linkage is approved by Statistics Finland's ethical committee and is performed using unique personal identification numbers available for all residents and then anonymised for research purposes. Statistics Finland has granted us permission to use the data for research (TK-53-1519-09).

We restrict our study population to individuals born between 1939–1950 who are aged 50–73 years in the period 2000–2012, are residents of Finland, and retire in 2005 or later.. We chose the period to cover a period before and after the 2005 pension reform and age-groups to cover ageing workers some of whom we could also observe both before and after the normative retirement age of 63 years. We exclude individuals from the treatment group who (a) are not living in private households; (b) have preceding periods of disability, unemployment or other routes out of employment before the old-age retirement or continue to work after receiving pension[5]; or (c) have not worked in the private sector. Restricting the study population to those living in private households in Finland is a natural choice, as the remaining population is mainly in institutional care and thus the de facto changes in the key elements of retirement transition are expected to be missing. The conditions listed in (b) are set because of the aim to study old-age retirement from work. Public sector workers have been ruled out because of different retirement rules dominated in that sector during the study period. In particular, the private sector workers share a common lowest statutory eligibility age for full old-age pensions, whereas in the public sector the rules are more ambiguous involving different retirement ages for different occupations.[6]

After the restrictions, our total sample includes 93,381 individuals and 1,148,465 observations over time. In total, 69,196 of the individuals retired during the study period (33,182 men and 36,014 women). Table 1 shows the descriptive statistics of our data for these people one year before their retirement. Furthermore, we isolate a treatment group that experience a clean transition from work to old age retirement according to our criteria. The treatment group includes 17,635 people; it is restricted to those entering retirement from employment.

3.2. Health outcome data

We use annual indicator variables [0,1] for three different health outcomes. For each of the study years, we assess the purchases of any antidepressant, defined as Anatomical Therapeutic Chemical code N06A, except tricyclic medication (N06AA). Hospital visits are coded on the basis of the International Classification of Diseases 10th Revision. Cardiovascular diseases are defined as codes I00-I99 and musculoskeletal diseases as M00-M99.

We believe our health measures are an improvement on most of the previous literature. First, as the measures are register-based they are not vulnerable to various self-report biases or missing information. Second, the hospital-based health indicators are based on information on over-night stays, thus avoiding many of the possible biases related to primary care utilisation. Third, all our indicators are likely to reflect more severe end of the assessed health problems. For example, even antidepressants are available only by prescription from a doctor after clinical assessment, and over-the-counter psychotropic medications are not available in Finland. Fourth, retirement related income barriers to care are likely to be small in Finland. A large part of the prescription medication costs are reimbursed directly regardless of income level and all medication costs are reimbursed after exceeding an out-of-pocket payment ceiling. As a last resort those on small incomes can cover costs of medicines and healthcare through social assistance (Hämäläinen et al. 2009). However, although the health problems we measure are significant, not everybody will seek treatment, and this tendency may be associated with relatively stable characteristics, such as sense of stigma or preference for self-sufficiency. To the extent that these factors are time invariant, they are controlled for in the individual fixed-effects design.

---

[5] In (b), we track whether a person has periods of disability or unemployment or has not worked more than six months in the year that preceded the retirement. We consider a person to continue working if there is at least one month of work in any year that followed retirement.

[6] While we do not directly analyse the health effects of retirement in the public sector, it is notable that the private sector data involves several occupational groups for which work characteristics are similar to the public sector work.

3.3. Socioeconomic status

Two indicators of socioeconomic status are used: occupational social class and individual taxable income. Occupational social class is based on the current main occupation before retirement and classified into seven categories: (1) upper non-manual, (2) lower non-manual independent work, (3) lower non-manual routine work, (4) specialised manual work, (5) non-specialised manual work, (6) manual work with unspecified degree of specification (7) farmer, (8) other entrepreneur, (9) student and (10) unidentified. We have excluded in our econometric study people from small unidentified classes and students and aggregated the data so that it separates non-manual and manual work only. The information on individual taxable income originate from the tax records and incorporated wages, capital income and taxable income transfers. Income is also measured in the year prior to retirement and is divided into quintiles and jointly calculated for men and women.

	Total obs.	Treatment obs.	Users of antideps.	Hospital periods (muscul.)	Hospital periods (cardio.)
Men	33,182	9,178	5.4%	3.8%	2.8%
Upper non-manual	5,749	1,887	6.5%	3.1%	2.4%
Lower non-manual independent	5,038	1,544	5.5%	3.3%	2.7%
Lower non-manual dependent	1,109	269	6.4%	3.1%	2.7%
Specialised manual	8,219	1,864	5.1%	4.2%	3.0%
Non-specialised manual	5,838	1,348	4.7%	3.6%	2.6%
Manual, unidentified spec.	791	0	5.7%	5.6%	1.6%
Farmer	2,398	678	4.2%	3.7%	2.9%
Other entrepreneur	3,541	1,588	4.9%	4.6%	3.7%
Student	339	0	10.6%	2.9%	1.5%
Other, unidentified	160	0	9.4%	6.3%	2.5%
Women	36,014	8,457	9.5%	2.1%	3.3%
Upper non-manual	4,964	1,205	10.7%	1.7%	2.9%
Lower non-manual independent	8,502	2,321	9.4%	2.3%	3.0%
Lower non-manual dependent	6,782	1,497	10.5%	2.0%	3.6%
Specialised manual	2,866	461	8.8%	2.1%	3.6%
Non-specialised manual	7,545	1,785	8.6%	2.3%	3.3%
Manual, unidentified spec.	639	0	11.3%	2.7%	3.1%
Farmer	1,895	342	6.9%	2.6%	4.1%
Other entrepreneur	2,263	846	9.4%	1.7%	3.1%
Student	403	0	13.6%	2.0%	4.2%
Other, unidentified	155	0	12.9%	4.5%	2.6%
Total	69,196	17,635	7.5%	2.9%	3.1%

Table 1. Descriptive statistics. Note: Total obs. is the total amount of people in the data measured one year before their retirement and separately reported for each socioeconomic group. The treatment obs. variable shows the number of retirements that are qualified as clean transformations from full-time work to retirement, again measured one year before retirement. Users of antideps is the share of people using antidepressants one year before retirement, while the hospital period variables indicate the share of people that were subjected to a treatment period in hospital in the year before their retirement.

3.4 Description of institutions

The Finnish public pension system consists of tax-financed basic pensions (a national pension and a guarantee pension) and statutory earnings-related pensions that employers and employees finance. Second-pillar occupational pensions and individual pensions play a marginal role. This improves the accuracy of our study, as the earliest eligibility age for full pension in the earnings-related pension scheme becomes a strong predictor for retirement.

The 2005 pension reform gradually removed popular early retirement schemes and adopted a flexible old-age retirement scheme for those aged 62–67 for both genders. It introduced early retirement at age 62, but with a heavy cut in pensions except for those who were long-term unemployed. Postponing withdrawal of pension was rewarded after age 63 in a way that was considered to be actuarial on average. In the public sector, some of the pre-existing occupational and personal retirement ages remained intact for the older generations.

The retirement peak, which before was seen at ages 60 and 65, moved to age 63; further, retirement directly from work to an old-age pension gradually became a rule rather than an exception[7]. Apart from the changes in the pension system rules, the falling disability incidence rate and employees' increasing average educational level also contributed to the higher employment rates of the retiring baby boom generations.

In order to generate a clean research frame, we chose to study retirement after 2005 for old-age pensions at ages 62–67 in the private sector. Since we focus on the direct transition from work to old-age pensions, we skip those individuals who were unemployed or disabled before their retirement, or who for other reasons were not working, since they do not see any change in the work/leisure margin and experience little change in their income. As an example, including disability pensioners – who in Finland are automatically transferred to old-age pension at their statutory full pension retirement age – in the old age pension after turning 63 would improve the first stage results of IV, but it would distort the estimates of the health effects of retiring from work. We also removed those who continue to work while receiving a pension from the data for the same reason. The rules for old-age pension remained the same during the research period of 2005–2012.

In addition to the details of the pension system, access to health care may also have important implications for the results. Almost all employees are eligible to use free occupational health care, which in many cases also provides medical treatment. This allows them to bypass the long queues at public health care centres. In addition to causing discomfort, waiting time may also reduce labour income. After retirement, this route ceases to exist. Therefore, any attempt to use number of visits to health care centres as an indicator of health is likely to bias the results, showing a weakening of health after retirement.

4. Research design

The analysis faces several challenges that arise above all from unobserved heterogeneity at the individual level. Individuals are different, for example, in terms of their health, environment and preferences toward retiring. Moreover, there are factors that may affect both health and the decision to retire, and the decision to retire may reflect changes in health rather than vice versa. Without controlling these factors, the statistical analysis may lead into biased or inefficient estimates. For example, if the decision to retire is made systematically after a negative change in health status, it is likely that the results will indicate that the

---

[7] Basically, using data on the reform of 2005 would have provided additional information on the individuals who retire because of becoming eligible. But the multitude of early retirement schemes and occupation and industry specific differences in normal retirement ages prevailing before the reform of suggest that individuals were faced with very different incentives to retire (occupation, sector, health, birth year, year when working career started, etc.) and were also likely to differ in terms of their unobserved characteristics. The retirement peaks in ages 60 and 65 were also small to be used in the analysis, and the ones who worked until age 65 were strongly selected.

retirement has a negative effect on health in the absence of controlling for the direction of the causality. That may conceal the true, positive effects of the retirement, if they exist.

We use an individual fixed-effects IV design. The idea is, as is common in the recent literature, to use the lowest statutory eligibility age for full old-age pensions as an instrument. It is a good predictor of the true retirement age in Finland, and it is not likely that the statutory retirement age correlates with health. Therefore, we expect the statutory retirement age to provide a strong instrument for the true retirement year.

The discontinuity caused by the statutory age act as instruments for individuals' employment status in a 2SLS model, with age as the continuous variable determining the discontinuities (Angrist and Pischke 2009; Imbens and Wooldridge 2009). In the formulation of the framework, we follow Heller-Sahlgren (2017). Let us denote the retirement with the variable r_{it} – it receives a value of 1 if the person has retired – otherwise, it receives a value of 0. Then, P, the probability of the treatment ($r_i = 1$), can be expressed as a piecewise continuous function of the age above and below the threshold, sp_i (in our case 63 years):

$$P(r_i = 1 | age_i) = \begin{cases} f_1(age_i) \text{ if } age_i \geq sp_i \\ f_0(age_i) \text{ if } age_i < sp_i, \end{cases} \text{where } f_1(age_i) \neq f_0(age_i), \quad (1)$$

We expect that $f_1(age_i) > f_0(age_i)$, that is, that the probability of retirement increases after the person has reached the statutory eligibility age. We typically assume f to be either a linear or quadratic function of age, depending on the size of the estimation window.

In quadratic form, our estimation equations are

$$r_{it-1} = \beta_0 + \beta_1 \overline{sp}_{it-1} + \beta_2 age_{it} + \beta_3 age_{it}^2 + \beta_4 age_{it} \overline{sp}_{it} + \beta_5 age_{it}^2 \overline{sp}_{it} + \gamma_i + \omega_t + \epsilon_{it} \quad (2),$$

and

$$health_{it} = \alpha_0 + \alpha_1 \hat{r}_{it-1} + \alpha_2 age_{it} + \alpha_3 age_{it}^2 + \alpha_4 age_{it} \overline{sp}_{it} + \alpha_5 age_{it}^2 \overline{sp}_{it} + \gamma_i + \omega_t + \epsilon_{it}, \quad (3),$$

where $\overline{sp}_{it}$ is an indicator variable that receives a value of 1 if $age_{it} \geq sp_i$ and receives a value of 0 when $age_{it} < sp_i$. When the estimation is done locally at the window (statutory retirement age – 1, statutory retirement age + 1), we instead use only a single linear aging trend around the discontinuity generated by the legislation ($\beta_3, \beta_4, \beta_5, \alpha_3, \alpha_4, \alpha_5 = 0$).

Similar to Heller-Sahlgren's (2017) notation, $\hat{r}_{it-1}$ is the prediction of r_{it-1} from the first stage, with $\overline{sp}_{it-1}$ used as an excluded instrument.[8] Thus, our baseline model estimates the average effect of retirement to on the different health variables at in the years that follows the year of retirement and after that.[9] We use the year- and person-fixed effects, ω_t and γ_i, respectively, during the estimation; we also use a quadratic trend effect of aging on the health outcome. Furthermore, it should be acknowledged that the person-specific heterogeneity in the effect of the retirement is likely to cause dependence in the error terms at the individual level; thus, the estimator should be chosen in a manner that is consistent with the clusters in the errors, ϵ_{it}.

---

[8] Our instrument is likely to fulfil the relevance condition (retirement concentrates strongly at age 63), exclusion condition (the statutory lowest eligibility age influences health only through retirement) and monotonicity condition (the statutory retirement age only increases actual retirement). The results provide the local average treatment effect (i.e. the impact of retirement on health for those who retire from work because of reaching age 63).

[9] This is because the amount of work that a person has done during his or her retirement year can vary, and it is not clearly specified in our data. Our data show how many working months occur during a person's final year, but some of them may constitute paid vacations that people often save to use at the end of their working careers. In our robustness analysis, we, however, show that our main results remain similar when we analyse the contemporaneous effects of retirement, that is, we use the prediction of r_{it} from the first stage, with $\overline{sp}_{it}$ used as an excluded instrument.

The estimation design bears resemblance to the standard fuzzy regression discontinuity design (RDD). However, the focus is on the variation within individuals across time rather than the variation between individuals (see Heller-Sahlgren 2017; Eibich 2015; Lemieux and Milligan 2008; Petterson-Lidbom 2012). Individual-level fixed effects are included in the model, and as a result, the identification assumptions are different. A traditional fuzzy RDD would hinge on the assumption that people who are on different sides of, but close to, the statutory age limit only differ in terms of the probability of being retired, once the flexibly for the direct impact of age is controlled for. Instead, the individual fixed-effects IV estimator hinges on the assumption that merely crossing the threshold serving as an instrument does not impact an individual's health around the threshold apart from via retirement (Heller-Sahlgren 2017).

β_1 represents the difference in the probability of treatment between the control group and the treatment group. The difference is often called the first stage; it gives an estimate of the impact of reaching the retirement age on the fraction of treated individuals. Typically, the research generates a LATE – that is

$$LATE = \frac{\alpha_1}{\beta_1}. \ (4)$$

Despite its conceptual appeal, the IV method is not without problems. First, under-identification of the instruments may occur, which means that some or all the instruments are irrelevant, as they are not sufficient to identify the relationship between the endogenous regressors and the explained variable. Weak identification arises when the excluded instruments are correlated with the endogenous regressors, but only weakly. Estimators can perform poorly when instruments are weak (see Stock and Yogo [2005] for further discussion).[10]

However, even when the standard tests reject weakness of the instrument, the external validity of the LATE may be violated when β_1 is small and there is evidence against the homogeneity of the expected marginal treatment effects in the population. In this case, the group is selected and the selection matters for the outcome.

Our analysis suggests that one should be very careful when interpreting the results of studies that use the statutory retirement age as an instrument. That is because in most of the study settings continuing to work until the statutory retirement age may not be common and this may lead to a partial selection of exposure to retirement (of treatment). Some people choose to retire before the retirement age, and others who continue to work forgo the possibility of retiring at the statutory age. Thus, the research design is quasi-experimental, with two-sided noncompliance (Brinch et al. 2017). In the general terminology of Angrist, Imbens, and Rubin (1996), the treated (retired) persons consists of compliers whose behavior is affected by the binary instrument at hand (crossing the age threshold), but also always-takers who are treated (retired) irrespective of whether the instrument is switched off or on (the age threshold is crossed). The untreated (not retired) persons are likewise composed of compliers and never-takers, where the latter group avoids treatment even when the instrument is switched on, I.e. the age threshold is crossed.

In the current retirement literature, there is typically only a small increase in the probability of entering retirement at reaching the statutory age, i.e. the share of those who retire at 63 because of the instrument. The typical values of the first stage estimate of the increase of probability due to reaching the threshold age

---

[10] In the estimations and statistical testing, we use the *xtivreg2* module for Stata by Schaffer (2010). The under-identification test is an LM test of whether the rank of the matrix of reduced form coefficients is smaller than the dimensionality of the problem. Under the null condition, the statistic is distributed as chi-squared, and a rejection of the null indicates that the matrix is of full-column rank (i.e. the model is identified, and the rejection is based on the Kleibergen-Paap [2006] rk statistic). In addition, we use the Kleibergen-Paap Wald rk F-statistic with the degrees of freedom adjustment for the rk statistic following the standard small-sample adjustment for cluster–robust standard errors.

are well below 0.5. While the IV estimation provides an unbiased estimate of the LATE of the treated, the larger the problem of noncompliance, the harder it is to know how well the result generalises to the larger population. When the share of those who react to the instrument is only a small fraction of the total population, it is a particular group that may be especially prone to either the positive or negative health effects of retirement.

The novelty of this paper is that we test the homogeneity of the treatment effect by resorting to Brinch et al.'s (2017) proposed method. The test of external validity combines several test statistics that compare the outcomes of intervened and non-intervened people in groups that are similar in their treatment status. Intuitively, if the outcomes differ in these groups, the population has heterogeneous marginal treatment effects as a response to the intervention (the eligibility age. This would invalidate the generality of the LATE outside the complier group.

We focus on investigating the generality of our results within a narrow age window surrounding the earliest eligibility age. Generalisations of the results even in the closest age interval of 62–64 is relevant because the majority of the Finnish private sector employees retire from work at this age.

Taking the [62,64] age window as a reference, our first test statistic measures the difference in the average untreated health outcomes between compliers (at age 62, just below the eligibility age and those who retire at 63) and those who have turned 63 but will retire at 64. Because it compares the untreated outcomes, the test statistic can be referred to as the untreated outcome test statistic. If the statistic is different from zero, it is a sign of selection heterogeneity: There is a difference in the health outcomes of the *untreated* that is associated with the instrument-induced selection of treatment. Those who have preferences against retirement, indicated by postponing it, are different from the compliers.

The analogous treated outcome test statistic measures the difference in average outcomes of those aged 62 and retired in the same year, and those aged 63 and retired at 63). If the statistic is different from zero, it is again a sign of selection heterogeneity. The instrument-induced selection, as reflected by the change in treatment probability, shows up as a difference in the outcomes of the *treated (i.e. retired)*. Those who have preferences toward taking the treatment (early) are different from those retired at 63.

Formally, we define the untreated and treated test statistics, Δ_0 and Δ_1, respectively, as follows:

$$\Delta_0 = E[health_{it}|age_{it} = 63 \text{ ,retires at 64}] - E[health_{it}|age_{it} = 62 \text{ , retires at 63}]$$

$$= \Delta age + \Delta_0 E[\gamma_i] + \Delta_0 E[\epsilon_{it}] \quad (5)$$

$$\Delta_1 = E[health_{it}|age_{it} = 63 \text{ ,retires at 63}] - E[health_{it}|age_{it} = 62 \text{ ,retires at 62}]$$

$$= \Delta age + \Delta_1 E[\gamma_i] + \Delta_1 E[\epsilon_{it}] \quad (6)$$

We trivially decompose the test statistics into several effects: (1) the mean effect of aging, Δage, which we define -without any loss of generality - to be common within any individual group for which the estimation is conducted; (2) the mean time-invariant differences in the health outcomes between the groups, $\Delta E[\gamma_i]$; and (3) the mean differences in the model's idiosyncratic errors that capture the heterogeneity of the changes in the health status of treated and untreated. We note that the test statistics ultimately boils down to analysing the heterogeneity in term (3) while we include term (1) merely to reflect the structure of the model that uses discontinuity caused by the statutory age in the age profile of health as means to analyse the effect of retirement.

Our test statistic of external validity compares the size of the treated and untreated test statistics, $\Delta_1 - \Delta_0$, and it rejects the null hypothesis of treatment effect homogeneity if the untreated outcome test statistic is

not equal to the treated outcome test statistic.[11] Those who have preferences toward or against taking the treatment are different in their health outcomes. The treatment effect homogeneity is also a test for external validity because the LATE can only be externally valid if the treatment effect is homogeneous. If the treatment effect is homogeneous, then the treated outcome test and the untreated outcome test only reflect selection. If instead, the two test statistics are different, there is selection to the treatment *and* the treatment effect is not homogeneous. Then, the LATE is not externally valid.

Formally, the test statistics can be written as:

$$\Delta_1 - \Delta_0 = (\Delta_1 age - \Delta_0 age) + (\Delta_1 E[\gamma_i] - \Delta_0 E[\gamma_i]) + (\Delta_1 E[\epsilon_{it}] - \Delta_0 E[\epsilon_{it}]), \qquad (7),$$

where the first term (the within-group age profile of health) is expected to be zero, the second term captures the differences in the time-invariant health outcomes, and the last term captures the time-variant heterogeneity. If the third term is not 0, the test statistics indicate a rejection of the homogeneous treatment effects. The test can be employed using standard t-testing. Furthermore, the hypothesis that the test statistics for several groups are jointly not different from zero can be studied using the F-statistics.

To build further intuition behind the test statistics of external validity, we note that it essentially measures the effect of having a stronger instrument on the LATE. The stronger instrument increases the number of compliers and decreases the number of both the never takers and always takers. Brinch et al. (2017) shows that under the standard IV assumption of conditional independence and monotonicity, the marginal effect of increasing the number of compliers on the LATE due to a decrease in the amount of never takers is proportional to $-\Delta_0$. Similarly, the marginal effect of increasing the number of compliers on the LATE due to a decrease in the amount of always takers is proportional to Δ_1. In particular, the conditional independence of the instrument of the idiosyncratic error terms, ϵ_{it}, implies that the strengthening of the instrument does not change the third term in Equation 7; therefore, the extrapolation of the test statistics on the larger complier group is valid. Under the (local) linearity of the effects, the proportionalities are the same; therefore, it follows from $\Delta_1 - \Delta_0 = 0$ that the LATE remains the same even when the number of compliers increases.[12]

Access to the test statistics allows us to test different ways of avoiding heterogeneity of the effects in the quasi-experimental design. As compared to the earlier work by Heller-Sahlgren (2017) that uses waves of survey data with irregular frequency, our annual data provide more detailed information about the date of a person's retirement and his or her health outcomes. This allows us to construct the test statistic, as we can isolate the exact year of retirement as well as the health outcomes prior, at the time and after the retirement. Furthermore, we have access to a long time series of personal health outcomes that allows us to properly control for person-level fixed effects. Furthermore, we can tackle heterogeneity by controlling the individuals' observable characteristics and by studying the effects of the retirement separately for different groups. We also consider matching in order to balance the data with respect to observable characteristics.

In addition, we aim to avoid other common pitfalls of the previous IV studies based on regression discontinuity design. We allow the impact of age to differ on each side of the retirement discontinuities used

---

[11] This is a sign of treatment heterogeneity under the assumption that the functions that specify how treated and untreated outcomes vary with the fraction of the treated *P* are linear.

[12] We acknowledge that the MTE might be non-linear in some occasions. We note that Brinch et al. (2017) also proposes a version of the test statistic that generalizes to monotonic MTE but at the expense of losing statistical power. We motivate the use of more powerful test by the fact that we mainly apply the test in a small age window around the eligibility age in which case the linear model is likely to provide us with a reasonable approximation of the MTEs.

as instruments (Lee and Lemieux 2010), and we study the sensitivity of the findings by narrowing the range of the data analysed around the retirement discontinuities (Angrist and Pischke 2015) in order to explore the non-linear effects of statutory retirement age or the results' sensitivity to the specific data range. Instead of using survey data, we employ yearly register-based panel data to identify both the date of retirement and changes in health outcomes.

We pay special attention to the definition of retirement and the statutory retirement age. In most previous research instruments are constructed from both regular and early retirement ages, which neglects potential self-selection into jobs where individuals are more likely to be able to retire early (Heller-Sahlgren 2017). Furthermore, working intensity before retirement varies. In terms of the institutional setup, it is often unclear which statutory retirement age threshold applies to which segment of the population and at what age. Asserting discontinuities at retirement ages without knowing whether any clear incentives for retirement can induce additional measurement errors.

Finally, it is worth noticing that many of the recent studies have used pension reforms to isolate the group that retires because of becoming eligible to pension. Those who are not affected by the reform (because of age, occupation, gender, etc.) are in the control group. This setup may help to overcome some of the problems concerning the identification, as the discontinuities can generate clear incentives for retirement across narrow age windows. Another useful aspect of the reforms is that one can observe age profiles of health outcomes prior to the reform. On the other hand, the reform approach is not without problems. It still compares health outcomes of different persons over time, and thus the risk of selection remains especially when the take-up of utilizing the discontinuities is typically relatively low. Moreover, the reform approach builds on the comparison of few cohorts at the vicinity of the discontinuity. The reform also generates other outcomes, such as change in incomes, that can influence health.

In our case, separation of the compliers from the control group is based on their retirement choices, when the earliest full pension retirement age is reached. With this method, it is more likely that the group that retires at age 63 includes also some who retire because of other reasons than becoming eligible. On the other hand, using the fixed retirement age allows us to follow the retirement behavior of eight successive birth cohorts and draw conclusions on wider data. Moreover, even with a single threshold, it is possible to use "Placebo" approach to study whether there are general health changes around the age threshold for persons who do not retire at that age.

5. Results

We start by considering the aggregate effect of retirement on health. Table 2 shows the fixed effects panel regression estimates for the effect of retirement on our health variables after controlling for the individual- and year-level effects. The results suggest that the correlation that arises from the individual-level variation in the retirement status and the health outcomes relatively low. It appears that retirement does to some extent positively affect the health status of the treatment group relative to the health status of the same-aged control group for the same year, but the results may be biased due to endogeneity problems.

		Use of antidepressants	Hospital periods (cardiovascular)	Hospital periods (musculoskeletal)
Retirement		-0.0052***	-0.0030**	-0.0064*'
		(0.0018)	(0.000838)	(0.000835)
Observations		210,465	210,465	210,465
Number of individuals		17,635	17,635	17,635

Table 2. Fixed-effects panel estimation of the aggregate effect of retirement on health. The coefficient of the retirement variable denotes the average effect of retirement to the health variable at the year that follows the retirement and after that. In the results, -

0.01 denotes a 1 percentage point decrease in the risk of using antidepressants or facing a hospital period when retired. Robust standard errors for heterogeneity and person-level clusters in the error terms in parentheses: ***p < 0.01, **p < 0.05, *p < 0.1. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.

We then repeat the estimation using the IV methodology in Equations 1 to 3. The results in Table 3 suggest that especially the impact on the use of antidepressant becomes statistically and economically significant. The results reported on the line 'LATE' suggest that retirement decreases the probability of using antidepressants by -3.05 percentage points. We find a smaller but still statistically significant effect on the risk of facing a hospital period due to a musculoskeletal disease or a cardiovascular disease; another finding that only becomes apparent after using the IV estimation.

While we do not explicitly model the determinants of the difference between methodologies, the likeliest reason is the reverse causality. The decision to retire is affected by negative health shocks; therefore, in absence of controlling for it, the decreasing effect of retirement on the use of antidepressants per se is muted. In terms of identification, it should be noted that we control for quadratic aging trends for each health variable; we estimate the effect of aging to be almost linear. We also find that the under-identification and the weak identification tests fail to reject the instrument's relevance and robustness.[13]

To further characterise the results, we report the reduced-form estimates of the effect of turning 63. The reduced-form model uses the same controls as the IV estimation. We find that turning 63 reduces the risk of using antidepressants roughly by 1 percentage point and the risk of a hospital period due to musculoskeletal disease by 0.5 percentage points. The reduced form is the IV estimate as weighted by the first stage, and it can be used to compare the overall feasibility of the model's predictions as relative to the corresponding changes in the average health outcomes in our sample of treated persons. We further illustrate the predictions of the reduced-form model and their uncertainty in Figures 4–6 in the Appendix. We show that the decline in the reduced-form estimates fits well with the actual changes in health outcomes in the sample, and the model also provides reasonable counterfactual dynamics for health outcomes in the absence of the effect of turning 63.

	Use of antidepressants	Hospital periods (cardiovascular)	Hospital periods (musculoskeletal)
Second stage			
LATE, α_1/β_1	**-0.0305*****	**-0.0080***	**-0.0163*****
SE	0.0050	0.0048	0.0053
First stage			
Pr jump, β_1	**0.3223*****	**0.3223*****	**0.3223*****
SE	0.0039	0.0039	0.0039
F-statistic	6877	6877	6877
Reduced form			
Effect of turning 63, α_1	**-0.0098*****	-0.0026	**-0.0053*****
SE	0.0016	0.0016	0.0017

---

[13] We find that the p-values for rejecting the under-identification are very close to 0, while the Kleibergen-Paap Wald rk F-statistic concerning the weakness of the instruments is of an order of magnitude higher than the benchmark critical values that Stock and Yogo (2005) provide. While in most cases, Stock and Yogo's (2005) critical values are not directly applicable due to our specification that does not impose i.i.d. errors, the large size of the F-statistics as compared to the standard critical values, and the rejection of weak identification in the first stage results indicates the lack of weak identification.

N (obs)	210,465	210,465	210,465
N (people)	17,635	17,635	17,635

Table 3. The benchmark IV estimation of the aggregate effect of retirement on health. In the results, -0.01 denotes a 1 percentage point decrease in the risk of using antidepressants or facing a hospital period when retired. Robust standard errors for heterogeneity and person-level clusters in the error terms in parentheses: ***p < 0.01, **p < 0.05, *p < 0.1. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.

We then analyse the effect more closely in Table 4 by repeating the IV estimation separately for different groups while still controlling for the aging effects and the appropriateness of our instruments (Figures 8-10 show the reduced-form fit of these models to the data in case of different occupational social classes and sexes). It should be noted that in each alternative specification, we fail to show that the instruments would be weak or irrelevant, whereas the aging trends remain close to linear. The first stage results are available in Table A2 in the Appendix.

Our results show a reduction in antidepressant use for women in almost all occupations, whereas for men the effect only is significant for manual workers and farmers, while the effects are much smaller for the other occupational categories. We also analysed the influence of retirement on the risk of requiring hospital treatment to cardiovascular diseases. The results are much weaker; we find that retirement only decreases this risk for individuals in a few of the socioeconomic groups. The beneficial effect of retirement on individuals with musculoskeletal problems is evident in non-manual work and in dependent non-manual work of both sexes.

		Second stage					
		Use of antidepressants		Hospital periods (cardiovascular)		Hospital periods (musculoskeletal)	
		LATE, α_1/β_1	SE	LATE, α_1/β_1	SE	LATE, α_1/β_1	SE
MEN	Non-manual	-0.0024	0.0103	0.0098	0.0115	**-0.0244****	0.0107
	Manual	**-0.0222*****	0.008	-0.0138	0.0105	-0.0155	0.0105
	Farmer	**-0.0731***	0.0375	**-0.0983***	0.0503	0.0244	0.0433
	(Other) entrepreneur	-0.0074	0.0126	-0.019	0.0177	0.0004	0.0138
WOMEN	Non-manual	**-0.0362*****	0.0119	-0.0051	0.0083	**-0.0197***	0.0112
	Manual	**-0.0685*****	0.0148	-0.0075	0.011	**-0.0292***	0.016
	Farmer	-0.0887	0.0544	-0.0248	0.0475	0.0623	0.0648
	(Other) entrepreneur	**-0.0472***	0.0244	0.0031	0.0198	0.005	0.0259
MEN	Single	-0.019	0.0168	0.0203	0.021	-0.0223	0.0179
	Non-single	**-0.013****	0.0061	**-0.0156****	0.0077	**-0.0137***	0.0073
WOMEN	Single	**-0.0725*****	0.0208	-0.0147	0.013	-0.0165	0.0178
	Non-single	**-0.0383*****	0.0092	-0.0034	0.0072	**-0.0188***	0.0101
MEN	Income ≥ 75%	-0.0002	0.0189	**-0.0481***	0.0258	-0.0257	0.0183
	75% > income ≥ 50%	-0.0105	0.019	-0.0357	0.0245	-0.0182	0.0222
	50% > income ≥ 25%	**-0.034*****	0.0106	0.0055	0.0132	-0.0019	0.0127
	25% > income	-0.0044	0.0078	-0.0086	0.0096	**-0.0211****	0.0091
WOMEN	Income ≥ 75%	-0.038	0.0304	0.0088	0.0236	0.0393	0.028
	75% > income ≥ 50%	**-0.053*****	0.0168	-0.0027	0.0112	-0.0002	0.0146
	50% > income ≥ 25%	**-0.0498*****	0.0127	-0.0102	0.0098	**-0.0368*****	0.0135
	25% > income	**-0.0461****	0.0193	-0.0045	0.013	-0.0189	0.0199

Table 4. The effect of retirement on health by group. Note: Robust standard errors for heterogeneity and person-level clusters in the error terms in parentheses: ***p < 0.01, **p < 0.05, *p < 0.1. In the results, -0.01 denotes a 1 percentage point decrease in the risk of using antidepressants or facing a hospital period when retired. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.

We again illustrate the reduced-form estimates of the effect of turning 63 by showing the predictions of the reduced-form model and their uncertainty in Figures 4–6 in the Appendix; we also show the model's counterfactual dynamics for the health outcomes in the absence of turning 63. Table 4 indicates which of the differences between the counterfactual and the actual dynamics are statistically significant.

In order to better understand the results, we have repeated the estimations by decomposing the effect of retirement by personalising the effect of retirement results, including his or her income and family status (see Table 4). In particular, we report the results based on the division of the effect into income quantiles. They show that the women's antidepressant use decreases particularly strongly in the lowest income quartiles, whereas for men, the effect is strongest at the lower-middle-income quartile. In case of the result for women, it should be noted that the results may partly be weakened by heterogeneity of the treatment effect, as discussed in Section 6. Finally, the risk of facing a hospital period due to a musculoskeletal disease lowers especially in the lower-income quantiles.

We find that the results in the low end of the income distribution are intuitive. Ultimately, our approach builds on the economic incentives that the pension system generates for postponing retirement. The large effect on the low-income quantiles suggest that there may be a stronger trade-off between economic benefits of continuing at work and weaker health outcomes in these socioeconomic groups. That is, it may be that the large negative income effect of retirement prior to the lowest eligibility age has led some low-income retirees to wait for this age even as they face more adverse health effects of continuing at work than the retirees in other income groups.

When we study the retirement effects for alternative family status, the findings suggest that the positive health effects are more concentrated in people that are non-single. This holds true especially for men: non-single men gain in each of the analysed health fields. The effect is substantial also for single women, but the generalisability of this result, is brought into question by the possibility of selection in Section 6. While these results still reflect the positive effect of retirement, they also seem to interestingly suggest that there are potentially negative, compensating effects for some risk groups such as single men. For them, the key elements of retirement transition, such as increased leisure time and decrease in social activity, may result in more adverse health outcomes.

In order to further clarify the gender effects, we repeat our analysis for women and men in the specifications "The aggregate effect on men" and "The aggregate effect on women" in Table A3 in the Appendix. Especially in case of the use of antidepressants, the effects strike as stronger for women consistently with the evidence presented above.

All in all, our results support the positive mental effect of retirement found in the previous literature (Atalay and Barrett 2014, Zhu 2016, Oshio and Kan 2017). However, at the finer level there are some differences. In contrast to Belloni et al. (2016) who find the strongest results for men, we find that the mental effects are the strongest for women, albeit we also observe the positive effect for men. In case of men, we find that mental health improves for manual jobs in line with Mazzonna and Peracchi (2017), whereas for women the effect is more unified across socio-economic groups than their results would suggest.

In line with a few previous studies, our results show that the physical effects of retirement are health improving (Insler 2014; Atalay and Barrett 2014; Zhu 2016; Coe and Zamarro 2011). At more detailed level, Mazzonna and Peracchi (2017) find that health improves in physically demanding occupations, but in occupations where physical stress is low, retirement may weaken health. Our findings suggest that there are

also significant effects for the non-manual workers. Correspondingly with the finding by Eibich (2015) which indicates that lower-educated individuals benefit from better physical health, we find that the positive effect is especially strong in the lowest-income socioeconomic groups. We, however, observe the positive effect for both physical and mental health.

6. Robustness analysis

We next subject our empirical results to a battery of robustness tests. To summarize them, we find that the result that retirement decreases the probability of using antidepressants appears to be robust, while for the other health variables, the results are smaller and more mixed in terms of their statistical robustness.

6.1. Selection to treatment

In this subsection, we investigate whether selection could be a driving force behind our results. We first consider the external validity of our results and then we conduct more traditional testing based on the use of additional health variables and a matching algorithm.

The analysis is important because reaching the statutory retirement age leads to only partial selection of retirement. Some people choose to retire before the retirement age, and others who reach it forgo the possibility of retiring. In absence of homogeneity of the expected marginal treatment effects in the population, the group of 63-year-old persons who retire due to the effect of the instrument, is selected and the selection matters for the outcome, even when the standard tests of the validity of instrument are passed.

While we do not explicitly observe those persons, who do retire because of our eligibility age instrument, a useful first look on their characteristics is provided by observing persons who retire at 63 as compared to others.

The comparison is reported in Appendix Table A4. The table collects the shares of different socioeconomic classes in the two groups and compares their health outcomes one year before the retirement. At the aggregate level we find that the health outcomes are relatively similar, the most visible difference being the moderately higher frequency of antidepressant use in the group that retires at 63. A closer inspection suggests that the group that retires at 63 is typically younger at the year of retirement, and in this group the share of men is higher. There is more frequent use of antidepressants especially among women, and the distributions of the socio-economic groups are somewhat different. A notable feature is that the share of single women in the group that retires at age 63 is smaller than in the group that does not retire at age 63.

While the differences are not at the aggregate level very large, the observed heterogeneity at a more detailed level nevertheless raises the question how much does the selection matter for our main results. Thus, we next conduct statistical testing of the heterogeneity of the treatment effects.

	Use of antidepressants	Hospital periods (cardiovascular)	Hospital periods (musculoskeletal)
Treated test statistics, Δ_1	-0.0004	0.0094	0.0065
Untreated test statistics, Δ_0	-0.0043	0.0006	-0.0003
$\Delta_1 - \Delta_0$	0.0039	**0.0088***	0.0068
T-test statistic $\Delta_1 - \Delta_0$	0.6049	1.7952	1.1965
N (age 63, retires at 63)	6987	6987	6987
N (age 62, retires at 62)	1065	1065	1065
N (age 63, retires at 64)	4457	4457	4457
N (age 62, retires at 63)	6987	6987	6987

Table 5. Aggregate homogeneity testing with fixed-effects. Note: Test statistics are based on Equations 5 and 6.

When we control for the individual-level fixed effects and use the age window 62-64 (Equations 5 and 6), we find that the test statistics are typically of moderate size (Table 5). For the probability of using antidepressants and requiring hospital treatments due to musculoskeletal diseases, the results indicate that the heterogeneity is rather small and statistically insignificant. In terms of the probability of facing a hospital treatment due to a cardiovascular disease, the test statistic indicates the possibility of heterogeneous treatment at the 10% significance level.

Let us consider this result in more detail. The rejection of the homogeneous treatment hypothesis in the case of cardiovascular hospital treatments occurs because people aged 62 who retire at 62 are healthier than people aged 63 who retire at 63 when compared regarding their long-term average health ($\Delta_1 > 0$). Thus, there seems to be selection toward early retirement. In contrast, people aged 62 who retire at 63 are typically as healthy as people aged 63 who retire at 64 ($\Delta_0 \approx 0$). The findings provide evidence of selection toward retirement at the age 63, and because the treatment effects are heterogeneous among the different groups, the LATE health outcomes are likely to differ from the general population effects in the age window 62–64.

In terms of hospital treatments for musculoskeletal conditions, we find a similar pattern as with the hospital treatments for cardiovascular conditions, but the effect is smaller and insignificant. In terms of antidepressant use, we find weak signs that people who retire at 63 are typically less healthy at the age of 62 than people aged 63 and retire only at 64, that is, $\Delta_0 < 0$. That said, people aged 62 who retire at 62 are as healthy as people aged 63 who retire at 63 ($\Delta_1 \approx 0$). Again, the size and significance of the impact is small.

In order to further understand our results, we next measure the individual test statistics without controlling for the individual-level fixed effects but otherwise according to Equations 5 and 6. In case of all health variables, we find that the null hypothesis of homogeneous treatment effect is rejected at the 1% confidence interval. Second, we define the untreated and treated test statistics, Δ_0 and Δ_1, respectively, beyond the age interval 62-64:

$$\Delta_0 = E[health_{it}|age_{it} = 63, retires\ at\ age\ \geq\ 64] - E[health_{it}|age_{it} = 62, retires\ at\ 63]$$

$$= \Delta_0 age + \Delta_0 E[\gamma_i] + \Delta_0 E[\epsilon_{it}] \quad (8)$$

$$\Delta_1 = E[health_{it}|age_{it} = 63, retires\ at\ 63] - E[health_{it}|age_{it} = 62, retires\ at\ age\ \leq\ 62]$$

$$= \Delta_1 age + \Delta_1 E[\gamma_i] + \Delta_1 E[\epsilon_{it}] \quad (9)$$

We again reject homogeneity for all health outcomes.[14]

These results indicate that both the use of individual fixed effects to control for heterogeneity and limiting the observation window is important, and that once this has been done, the heterogeneity is of a moderate size. This finding is interesting, as it suggests that the results concerning health effects of retirement may not be generalizable if a standard RDD framework would be used without implementation of individual-specific fixed effects.[15]

---

[14] We abstract from the details of these results, while hey are available upon request from the authors.

[15] To further elaborate this point, we have also studied the characteristics of the compliers by comparing their health conditions to non-compliers in the spirit of Angrist and Pischke (2009, pp. 171). In particular, we study the ratio of the compliers' probability of the adverse health outcome and the overall probability (compliers + non-compliers). Arguably, the most natural choice in our context is to make the comparison for (untreated) persons aged 62. As the complier group, we focus on persons who will retire at 63 and as the non-complier group on those who will retire at 64. For the

In order to further illustrate the role of different sub-groups in the homogeneity testing, we repeat the analysis also for them. First, we find that the joint hypothesis of homogeneity (chi-squared distributed F-test statistics with degrees of freedom equal to the number of sub-groups), as measured separately for each decomposition of the aggregate effect, can only be rejected for the cardiovascular hospital periods.

While the aggregate results are in line with our previous analysis, the results of sub-group homogeneity testing in Table 6 shows that at the sub-group level there are a few interesting cases in which the homogeneity is rejected. These cases involve single women and low-income women in case of the use of antidepressants, and the lowest income men and entrepreneurs in case of cardiovascular diseases. One explanation of this pattern is that these groups may self-select into retirement based on their expected health benefits or particularly strong trade-offs between health and income. In any case, the impact of these groups on the aggregate results seem to be limited.

| | | Second stage and homogeneity tests | | | | | |
| | | Use of antidepressants | | Hospital periods (cardiovascular) | | Hospital periods (musculoskeletal) | |
		LATE, α_1/β_1	Homogeneity test (t-value)	LATE, α_1/β_1	Homogeneity test (t-value)	LATE, α_1/β_1	Homogeneity test (t-value)
MEN	Non-manual	-0.0024	0.34	0.0098	0.66	**-0.0244****	0.28
	Manual	**-0.0222*****	0.32	-0.0138	0.87	-0.0155	0.4
	Farmer	**-0.0731***	0.21	**-0.0983***	0.11	0.0244	1.56
	(Other) entrepreneur	-0.0074	0.21	-0.019	**2.35****	0.0004	0.08
WOMEN	Non-manual	**-0.0362*****	1.19	-0.0051	0.38	**-0.0197***	1.43
	Manual	**-0.0685*****	0.33	-0.0075	0.7	**-0.0292***	1.1
	Farmer	-0.0887	1.23	-0.0248	0.24	0.0623	1.14
	(Other) entrepreneur	**-0.0472***	0.12	0.0031	**4.4*****	0.005	0.84
MEN	Single	-0.019	0.21	0.0203	1.6	-0.0223	0.33
	Non-single	**-0.013****	0.07	**-0.0156****	1.09	**-0.0137***	0.07
WOMEN	Single	**-0.0725*****	**2.67*****	-0.0147	0.59	-0.0165	0.25
	Non-single	**-0.0383*****	0.66	-0.0034	0.63	**-0.0188***	1.47
MEN	Income ≥ 75%	-0.0002	0.29	**-0.0481***	0.18	-0.0257	0.43
	75% > income ≥ 50%	-0.0105	0.06	-0.0357	0.4	-0.0182	0.5
	50% > income ≥ 25%	**-0.034*****	0.19	0.0055	1.05	-0.0019	0.68
	25% > income	-0.0044	0.11	-0.0086	**2.43****	**-0.0211****	1.09
WOMEN	Income ≥ 75%	-0.038	0.93	0.0088	0.2	0.0393	0.12
	75% > income ≥ 50%	**-0.053*****	0.61	-0.0027	0.55	-0.0002	1.41
	50% > income ≥ 25%	**-0.0498*****	**2.01****	-0.0102	0.09	**-0.0368*****	0.27
	25% > income	**-0.0461****	0.75	-0.0045	0.88	-0.0189	1.03

Table 6. Homogeneity testing for individual groups with fixed effects. Note: Test homogeneity statistics are based on Equations 5 and 6 and they are calculated individually for each socio-economic group. The LATE estimates are based on Table 4.

The homogeneity testing provides only one source of information concerning the potential impact of selection. One way to further analyze our result is to study whether our results reflects selection through the short-term changes in health prior to retirement. To control for it, we include to our benchmark estimations

---

use of antidepressants, hospital periods (cardiovascular), and hospital periods (musculoskeletal) the ratios are 1.0269, 1.0597, and 1.0647, respectively. The results indicate that in the group of compliers the probability of the adverse health conditions tends to be marginally higher than in the group of non-compliers.

(in Table 3) the measures of other health outcomes one year prior to the retirement as additional control variables in the specification "Previous year's other health outcomes as additional controls" in Table A3. We find that our aggregate results are not sensitive to the inclusion of this additional information, again suggesting that the aggregate LATE findings are robust. We exclude the lagged dependent variable from the estimations, as its inclusion would lead into inconsistent estimation.[16]

In addition, we have also used coarsened exact matching algorithm to preprocess our data. This is in order to avoid the potentially confounding influence of pre-treatment control variables, and thus reducing imbalance between the treated and control groups. We balance our dataset in terms of gender, socio-economic groups, family status, and income groups for a sub-panel of persons that are retired at the age 64. When we repeat the benchmark analysis (in Table 3) in the specification "Estimation after matching", we find that the selection appears not to influence our findings.[17]

Finally, it should be noted that in the benchmark analysis we stratified data by focusing on the retirement transitions that we argue best represent the different elements of the retirement transitions. However, to ensure that our results are not due to the stratification, we repeat the analysis by considering retirements that may also involve unemployment before the retirement, or part-time continuation of work after the retirement in the specification "All retirements included" in Table A3. We find that the results are roughly half a percentage point larger in case of antidepressants use, while for other health variables the effect of using the full data either moderately increases (cardiovascular diseases) or has very small effect on the result (musculoskeletal diseases). All in all, we find that these moderate effects are consistent with our finding that selection does not have a major role in our aggregate results

6.2. Timing of the retirement effect
Our benchmark analysis shows results for a specific choice of the timing of the retirement effect. We estimate the average effect of retirement to the health variable at the year that follows the retirement and thereafter. This choice is natural as the amount of work that a person has done during his or her retirement year can vary, and it is not clearly specified in our data.

However, in order to analyze the robustness of our results in terms of timing, we also repeat our study by using the health outcomes that are measured at the year of retirement as the explained variable (instead of considering them at the year that follows the retirement in Table 3) in the specification "Contemporaneous effect of retirement" in Table A3. We find that the results remain very similar despite the change in the timing.[18]

---

[16] However, we have also conducted dynamic GMM panel estimations of the reduced-form model with several lags of the dependent variable as additional controls, while using the benchmark set of other control variables. We find that in each case the health effects are in line with our previous findings. In particular, we find that the effect on the antidepressant use is significantly negative and of similar size with our main results, while we do not find evidence of weak identification in the estimation context. The results are also significant for other health variables, but the instrument set may suffer from weakness.

[17] As in some cases we find evidence against generality of our LATE estimates, it is natural to ask, whether we could use alternative methods to make inferences on the ATE. For example, bivariate probit models could be used to estimate unconditional average causal effects under additional identifying assumptions (Angrist and Pischke, 2009). While our results based on the experimentation of the bivariate probit model indicate towards a positive effect, and this perspective would be highly interesting, we have not been able to include the large-scale panel structure of our data to the model, and thus further analysis in this direction remains to be done in the future work.

[18] A potential caveat of our analysis is that we cannot isolate whether the contemporaneous (t=0) effects are due to retirement or work in the year of retirement. This result suggest that the results do not change much due to the timing change, and thus we consider this not to be a major problem.

More generally, we illustrate the dynamics of the health effects over time by simultaneously estimating the health impacts of retirement for a broader time window. We estimate the effect each year in an interval that spans over a time period of 3 years before and 3 years after the retirement. Formally, we change Equations 2 and 3 to

$$r_{it+k} = \beta_0 + \sum_{k=-3}^{k=3} \beta_{1k} \overline{sp(k)}_{it} + \beta_2 age_{it} + \beta_3 age_{it}^2 + \gamma_i + \omega_t + \epsilon_{it} \qquad (10)$$

and

$$health_{it} = \alpha_0 + \sum_{k=-3}^{k=3} \alpha_{1k} \hat{r}_{it+k} + \alpha_2 age_{it} + \alpha_3 age_{it}^2 + \gamma_i + \omega_t + \epsilon_{it}, \quad (11)$$

after denoting $\overline{sp(k)}_{it}$ as an indicator variable that receives a value of 1 if $age_{it+k} \geq sp_i$ and receives a value of 0 when $age_{it+k} < sp_i$, and $\hat{r}_{it+k}$ are the predictions of r_{it+k} from the first stage, with $\overline{sp(-3)}_{it}$,…, $\overline{sp(3)}_{it}$ used as excluded instruments. [19]

In the dynamic model, we simplify the specification in order to strike a balance between the increased level of detail due to the dynamics and the level of detail of other features. Accordingly, we use a single, quadratic time path before and after the retirement for the full population, while we simultaneously estimate differentiated dynamic effects for male and female retirees. We concentrate on the retirements that occurred between the years 2005-2009 in order to focus on full paths of the transition dynamics within our time window. We include year-dummies and use clustered error terms, as before.

In Figure 2 and Figure 3 we show the percentage point deviation of the probabilities of the health events from the common time trend in the reduced-form model (collecting the coefficients of β_{1k}'s) and the IV model (collecting α_{1k}'s), correspondingly.

---

[19] We find that extending the estimation window, k, beyond 3, results in failure to reject the weakness test of the instruments. When $k = 3$, we find that the p-values for rejecting the underidentification are very close to 0. We find that the Kleibergen-Paap Wald rk F statistic related to the weakness of the instruments vary between 5 and 6 in the dynamic case, suggesting that they are close to the benchmark critical values provided by Stock and Yogo (2005). While it should be acknowledged that the critical values do not fully correspond to the current specification, the results raise the possibility of weakness of the instruments.
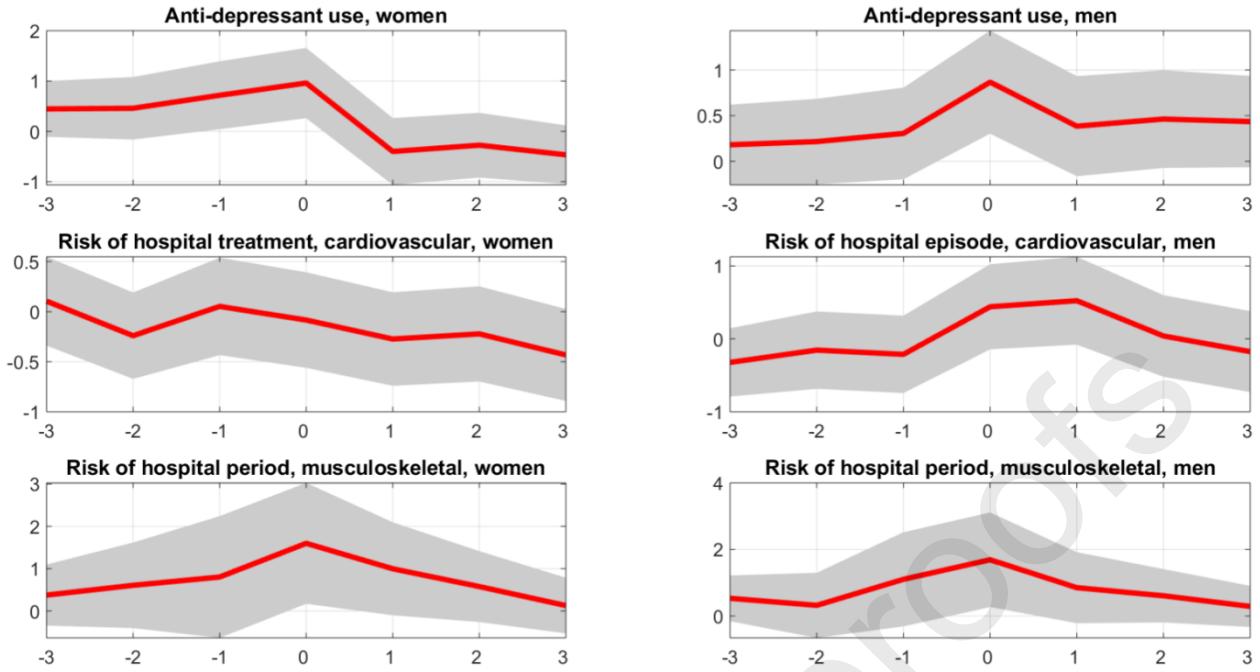
Figure 2. The reduced-form dynamics of the health effects over time. Note: We simultaneously estimate the reduced-form health impacts of retirement for each year in an interval that spans over the period 3 years before and after the retirement for women and men. The reported effects are β_{1k}'s in equation $health_{it} = \alpha_0 + \sum_{k=-3}^{k=3} \beta_{1k} \overline{sp(k)}_{it} + \alpha_2 age_{it} + \alpha_3 age_{it}^2 + \gamma_i + \omega_t + \epsilon_{it}$. Robust standard errors for heterogeneity and person-level clusters in the error terms, grey are is the 95 % confidence interval. In the results, y-axis denotes the % risk of using antidepressants or facing a hospital period. Other control variables: year- and individual-specific fixed effects and a quadratic aging trend.
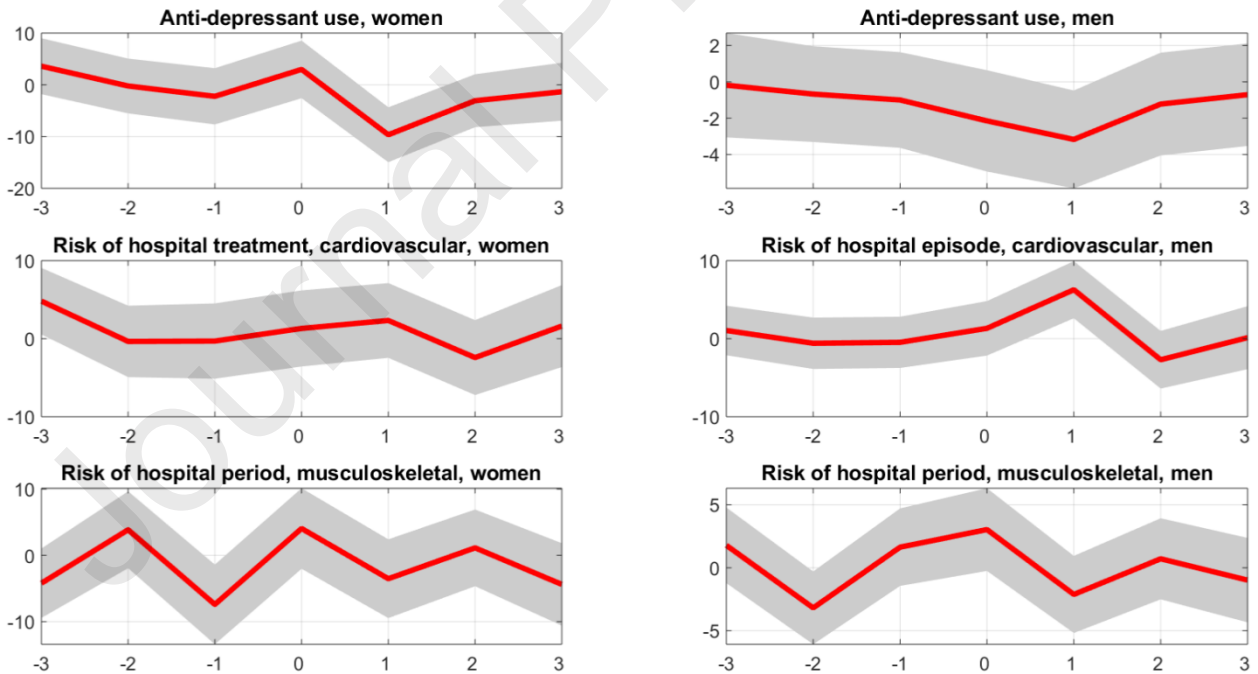

Figure 3. The IV estimation of the dynamics of the health effects of retirement. Note: We simultaneously estimate the IV health impacts of retirement for each year in an interval that spans over the period 3 years before and after the retirement for women and men. The reported effects are α_{1k}'s in Equation 11: $health_{it} = \alpha_0 + \sum_{k=-3}^{k=3} \alpha_{1k} \hat{r}_{it+k} + \alpha_2 age_{it} + \alpha_3 age_{it}^2 + \gamma_i + \omega_t + \epsilon_{it}$. Robust standard errors for heterogeneity and person-level clusters in the error terms, grey are is the 95 % confidence interval. In the results, y-axis denotes the % risk of using antidepressants or facing a hospital period. Other control variables: year- and industry-specific fixed effects and a quadratic aging trend.

We find that the estimates of the effect for a single year include a considerable amount of uncertainty that may reflect both the difficulty of making exact inferences on the dynamics and the heterogeneous effects across the socio-economic groups. Nevertheless, some dynamic features strike us as statistically and economically significant.

First, the women's risk of using antidepressants decreases sharply after the retirement as relative to the control group; an effect that is somewhat reversed in the later years according to the IV -estimates. The magnitude of the effect based on the IV estimates is substantial, roughly 10 percentage points, for the persons who retire as a result of reaching the age of 63, whereas the reduced-form estimate is only around 1 percentage points. This number is feasible when compared to the population level overall decline at the age 63, while findings suggest that the first stage is even smaller than in the baseline estimations, thus limiting even further the group of persons who retires as a result of crossing the threshold age. For men, the pattern is similar, but weaker, consistently with our previous results.

For other health variables, the effects are more diffused and slower to emerge. In case of musculoskeletal diseases, the declining effect is still relatively strongly linked to the retirement, whereas for cardiovascular diseases the effect starts only a few years after the retirement. Interestingly, we find that among men retirement may even have an initial, increasing effect on the possibility of cardiovascular diseases.

Interestingly, in contrast to Heller-Sahlgren (2017), we do not find any negative longer-term effects of retirement, albeit that our maximum length of the effect is only 3 years after the retirement.

6.3. Alternative aging profiles and "Placebo" effects

Correct characterization of the underlying aging trend is important part of our analysis. The discontinuity caused by the statutory age acts as instruments for individuals' retirement status in a 2SLS model, but only if it is properly identified by using age as the continuous variable determining the discontinuities. We next study in more detail the robustness and validity of this approach.

In the baseline analysis, we use the panel structure of our data, and estimate the impact of aging by using quadratic aging trends and all available data of the treatment group. This is somewhat different from the standard RDD framework in which the estimation window in terms of the running variable is chosen according to an optimization procedure. Moreover, a linear model is typically used to control for the effect of changes in the running variable in the vicinity of the cutoff.

It is notable that we cannot follow this approach because the running variable is not continuous, but rather it is characterized by few discrete mass points constituted by the yearly observations. In absence of the possibility to optimize the size of the estimation window, we instead repeat the analysis for different sizes of the estimation window to analyse the sensitivity of our results to it. We first conduct the analysis in the smallest age window 62–64 around the statutory retirement age while controlling for the age effects with a linear time trend. The results are reported in Table in the row "Estimation at the age window [62,64]". Moreover, we repeat the analysis for a moderately larger age window 61-65 in the row "Estimation at the age window [61,65]".

We find that the main results are robust to changes in the age window, while the effect of retirement on the use of antidepressants is moderately decreased when the smaller estimation window is used.

Another useful way to check the validity of our results is to test whether in the age 63 there is a discontinuity in the health profiles of persons who do not retire near this age. If such a discontinuity is observed, it would invalidate our analysis, as it would suggest that there is a direct health effect of turning 63, and thus other factors than retirement would be likely to explain our results. If instead, the effect is missing, that would provide additional confirmation to the validity of our approach. Accordingly, we conduct this "Placebo" test

by repeating the reduced-form baseline analysis for persons who retire before they turn 61, or after they have turned 65.

The results of the "Placebo" estimation are reported in Table 7. In support of our previous analysis, we find only small and insignificant effects of turning 63 on health for persons who do not retire at or near this age, when the direct effect of age is controlled similarly to our baseline analysis.

	Use of antidepressants	Hospital periods (cardiovascular)	Hospital periods (musculoskeletal)
Reduced form			
Effect of turning 63, α_1	-0.0010	0.0035	0.0016
SE	0.0031	0.0027	0.0032
N (obs.)	58482	58482	58482
N (persons)	4907	4907	4907

Table 7. Placebo testing. Note: The table shows the benchmark reduced-form estimates when estimation is based on the health data of persons who retire before turning 61, or after turning 65. Robust standard errors for heterogeneity and person-level clusters in the error terms in parentheses: ***p < 0.01, **p < 0.05, *p < 0.1. In the results, -0.01 denotes a 1 percentage point decrease in the risk of using antidepressants or facing a hospital period when retired. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.

7. Conclusions

Our objective in this study is to determine whether retirement from work to old-age pension influences various health indicators, and whether this impact differs by sex, occupational social class and income. A unique aspect of the data is that they are fully based on administrative registers, have a large sample size and no self-report or selective participation bias.

We find that retirement moderately decreases the use of antidepressants, especially for women. The effect is sharp, while it is somewhat reversed in the later years. The beneficial effects of retirement on the cardiovascular and musculoskeletal conditions are smaller and more diffused. Our results show a reduction in antidepressant use for women in almost all occupations, whereas for men the effect is significant only for manual workers and farmers, while the effects are much smaller for the other occupational categories. Moreover, we find stronger declines in the anti-depressant use among men and women who retire from low- and mid-income work, as well as more robust decrease in the risk of cardiovascular diseases for men in the highest income jobs and for non-single men.

We use the individual fixed-effects instrumental variable method to ensure causality. The lowest eligibility age for old-age pensions provides us with a strong instrument. The longitudinal register data allow us to follow the employment statuses and health of each person years before and after he or she retires and thereby precisely separate the transition from work to old-age retirement. The institutional structure of the Finnish pension scheme improves the coverage of the results. The first pillar earnings-related pension scheme dominates retirement behaviour, since the second pillar occupational pensions and third pillar individual pensions with their different retirement rules are marginal. Overall, we use new methods to evaluate

external validity – the results imply that the beneficial effects of retirement on the antidepressant use can be extended to apply to most Finns retiring from the private sector at ages 62–64.

Many of the previous studies, including some that use the Finnish data, did not detect an effect of old age retirement on mental health. Our study finds a positive influence of retirement on the purchase of antidepressants; this also applies to ages other than the earliest retirement age for full pension. Similar positive results have been found in different institutional settings, such as in Australia and Japan and the UK. These associations may reflect short-term honeymoon effects, and analyses of longer-term effects are called for. Our results highlight the importance of using high quality data and robust identification strategy in analysis, but also warns that there are potential selection problems in the standard quasi-experimental research design, at least in case of some health outcomes and socioeconomic groups.

In summary, even though the estimation results were largely conclusive in terms of statistical significance, the benefits of retirement on health were rather small and limited at the studied age window. On the one hand, our results imply that encouraging retirement may not be the most efficient policy to reduce the common health conditions among older workers. On the other hand, we argue that policy efforts to increase the retirement age only carry a relatively modest health penalty at the population level, albeit there are some socioeconomic groups that are affected more strongly.

Appendix

Additional tables

	Data	Country	Health variables	Retirement	heterogeneity	Main method	Instruments	Health effect of retirement
Atalay and Barret (2014)	Non-linked surveys	Australia	Self-assessed health, Diagnosed diseases	SR: Not in the labor force	Women	IV	Increase in standard retirement age	Positive
Behncke (2012)	Panel survey Elsa	UK	Diagnosed diseases	SR: Not in paid work, retired	No heterogeneity	Non-parametric matching, IV	Standard retirement age	Negative
Belloni et al. (2016)	Bi-annual panel survey SHARE	10 Countries	Self-assessed: mental health index Euro-D score	SR: Retired, not in labor force	Gender, region, blue/white collar	FE-IV	Early and standard retirement ages	Men positive, women no effects
Biro and Elek (2017)	Administrative panel	Hungary	Outpatient and inpatient expenditures, medicines	From work to full retirement	Women: sector, education	FE-IV	Increase in early retirement age	Lower expenditure
Bloemen, et al. (2017)	Administrative panel	The Netherlands	Mortality	From work to retirement	Male civil servants	FE-IV	Decrease in early retirement age	Lower mortality
Che and Li (2018)	Longitudinal survey	China	Self-assessed health	SR: Retired, not working	Male civil servants	FE-IV	Standard retirement age	Positive
Gorry et al. (2019)	Bi-annual survey HRS	USA	Self-assessed health, health indices, diagnosed diseases, utilization	SR: from work to retirement	Gender, educ., jobs, race, etc.	FE-IV	Eligibility ages	Positive for self-assessed health

Hagen (2018)	Administrative panel	Sweden	Drug prescriptions, hospitalizations, mortality	From work to retirement	Local government female workers	FE-IV	Increase in standard retirement age	No effects
Heller-Sahlgren (2017)	Bi-annual panel survey SHARE	10 Countries	Self-assessed: mental health index Euro-D score	SR: Retired, not in labor force	Gender, education, occupation	FE-IV	Early and standard retirement ages	Negative long-term effects
Hernaes, et al. (2013)	Administrative panel	Norway	Mortality	Leaving regular work	Gender, industry, occupation, education	IV	Decrease in early retirement ages	No effects
Hessel (2016)	Longitudinal survey EU-SILC	12 countries	Self-assessed: health, morbidity	SR: Retired, not employed	Gender, education, marital status	IV	Early and standard retirement ages	Positive for self-assessed health
Mazzonna et al. (2017)	Bi-annual panel survey SHARE	11 countries	Cognition, self-assessed: health, mental health index Euro-D score	SR: Retired	Physical demands of occupations	IV	Changes in early and standard retirement ages, distance	Negative for less demanding, positive for most demanding
Messe and Wolf (2019)	One-off survey	France	Self-assessed: health, chronic illnesses	Leaving last job before 60	Male workers in private sector	Probit, IV	Eligibility to early retirement	No effects from IV
Nielsen (2019)	Administrative panel	Denmark	Mortality, comorbidities, GP visits, hospitalizations	Receiving any pension, not employed	Gender, education, marital status	RD, IV	Decrease in early and standard retirement age	Early retirement: no effect on health, but lower utilization
Oshio and Kan (2017)	Longitudinal survey	Japan	Self-assessed: health, distress index score	From work to retirement	Gender	FE-IV	Distance to eligibility age	Positive effect on distress, for men positive for health
Shai (2018)	SHARE Israel, other surveys	Israel	Diagnoses, healthcare use and expenditure, self-assessed health	From work to retirement	Men	FE-IV	Increase in standard retirement age	Positive, especially for less educated
Zhang et al. 2018	Longitudinal biannual survey	China	Outpatient and inpatient use of services, costs, self-reported health	Retired, not working	Gender	Fuzzy RD	Standard retirement age	Increase in costs and utilization, negative for health

Table A1. Characteristics of recent studies. SR= self-reported.

		First stage				
		β_1	SE	F -test	N (obs)	N (people)
MEN	Non-manual	**0.3167***	0.0084	1405	44,091	3,700
	Manual	**0.3968***	0.0093	1815	38,214	3,212
	Farmer	**0.1912***	0.0196	95	8,080	678
	(Other) entrepreneur	**0.3545***	0.0134	698	18,921	1,588
WOMEN	Non-manual	**0.2952***	0.0072	1690	60,099	5,023
	Manual	**0.3285***	0.0104	989	26,866	2,246
	Farmer	**0.2042***	0.0264	60	4,069	342
	(Other) entrepreneur	**0.3108***	0.0182	291	10,125	846
MEN	Single	**0.3094***	0.014	486	16,640	2,144
	Non-single	**0.345***	0.006	3320	92,666	8,357
WOMEN	Single	**0.2591***	0.0099	684	28,873	3,215
	Non-single	**0.3157***	0.0067	2191	72,286	6,882
MEN	Income ≥ 75%	**0.3205***	0.0196	267	9,111	769
	75% > income ≥ 50%	**0.3166***	0.0171	341	11,383	958

Journal Pre-proofs

	50% > income ≥ 25%	**0.3614*****	0.0105	1187	29,508	2,478
	25% > income	**0.3459*****	0.0075	2138	58,709	4,919
WOMEN	Income ≥ 75%	**0.2734*****	0.0183	224	9,220	771
	75% > income ≥ 50%	**0.3469*****	0.0113	938	22,793	1,907
	50% > income ≥ 25%	**0.3089*****	0.0086	1296	41,740	3,487
	25% > income	**0.2649*****	0.0111	568	27,031	2,259

Table A2. First-stage results concerning the estimations reported in Table 4.

	Second stage						First stage				
	Use of antidepressants		Hospital periods (cardiovascular)		Hospital periods (musculoskeletal)						N (persons)
	LATE, α_1/β_1	SE	LATE, α_1/β_1	SE	LATE, α_1/β_1	SE	β_1	SE	F -test	N (obs)	
Previous year's (other) health outcomes as controls	**-0.0305*****	0.0050	**-0.008***	0.0048	**-0.0163*****	0.0053	**0.1693*****	0.0052	6879	210465	17635
Contemporaneous effect of retirement	**-0.0354*****	0.0076	**-0.0113***	0.0062	**-0.0195*****	0.0064	**0.2146*****	0.0051	5554	228126	17635
All retirements included	**-0.0346*****	0.0031	-0.0048	0.003	**-0.0163*****	0.003	**-0.0375*****	0.0022	26204	1054824	92731
The aggregate effect on men	**-0.0146****	0.0057	-0.0099	0.0071	**-0.0149****	0.0066	**0.2024*****	0.0074	3887	109306	9178
The aggregate effect on women	**-0.0497*****	0.0087	-0.0058	0.0062	**-0.0178****	0.0086	**0.1329*****	0.0074	3003	101159	8457
Estimation after matching	**-0.0394*****	0.0113	-0.0144	0.013	-0.0226	0.0139	**0.3124*****	0.004	951	188682	15799
Estimation at the age window [62,64]	**-0.0261*****	0.0073	-0.0096	0.0088	**-0.0195***	0.01	**0.2869*****	0.0045	4037	51053	17632
Estimation at the age window [61,65]	**-0.0202*****	0.0053	-0.0088	0.0059	**-0.0141****	0.0067	**0.323*****	0.0040	6399	81808	17634

*Table A3. Robustness analysis. IV estimation of the aggregate effect of retirement on health. Note: In the results, -0.01 denotes a 1 percentage point decrease in the risk of using antidepressants or facing a hospital period when retired. Robust standard errors for heterogeneity and person-level clusters in the error terms in parentheses: ***p < 0.01, **p < 0.05, *p < 0.1. Varying specifications.*

		Retires at 63	Others		Frequency of the condition, when retires at 63			Frequency of the condition, others		
		Share	Share	Average retirement age	Anti-depressants	Cardio-vascular	Musculo-skeletal	Anti-depressants	Cardio-vascular	Musculo-skeletal
AVERAGE		100 %	100 %	64.2	5.5 %	2.2 %	3.0 %	5.1 %	2.3 %	3.2 %
MEN		55.5 %	49.8 %	64.3	2.9 %	1.8 %	2.3 %	2.8 %	2.2 %	2.5 %
WOMEN		44.5 %	50.2 %	64.1	7.3 %	1.4 %	2.7 %	6.3 %	1.3 %	2.5 %
		100 %	100 %							
MEN	Non-manual	20.0 %	21.6 %	64.1	4.3 %	2.8 %	2.4 %	4.4 %	2.7 %	2.5 %
	Manual	21.7 %	15.9 %	64.3	3.2 %	2.5 %	3.5 %	2.7 %	3.5 %	3.7 %
	Farmer	3.2 %	4.2 %	65.0	3.1 %	2.7 %	1.8 %	1.8 %	2.4 %	3.1 %
	(Other) enterpreneur	10.6 %	8.0 %	64.7	3.4 %	2.2 %	3.9 %	2.7 %	3.8 %	2.8 %
WOMEN	Non-manual	25.5 %	30.5 %	63.9	8.1 %	1.8 %	2.9 %	7.9 %	1.5 %	3.4 %
	Manual	12.2 %	13.1 %	64.2	7.8 %	1.3 %	3.8 %	5.0 %	1.6 %	3.8 %
	Farmer	1.6 %	2.2 %	64.9	4.5 %	2.7 %	0.0 %	5.7 %	2.6 %	3.9 %
	(Other) enterpreneur	5.2 %	4.5 %	64.6	7.9 %	2.5 %	2.2 %	6.0 %	1.5 %	3.3 %
		100.0 %	100.0 %							
MEN	Single	8.7 %	8.3 %	64.5	4.4 %	2.8 %	2.0 %	4.5 %	4.0 %	2.9 %

	Non-single	46.8 %	41.5 %	64.3	3.5 %	2.5 %	3.3 %	3.1 %	2.9 %	3.0 %
WOMEN	Single	12.0 %	16.2 %	64.3	11.2 %	1.7 %	3.8 %	8.9 %	1.4 %	4.2 %
	Non-single	32.5 %	34.0 %	64.0	6.6 %	1.8 %	2.6 %	5.9 %	1.6 %	3.2 %
		100.0 %	100.0 %							
MEN	Income ≥ 75 %	4.8 %	4.1 %	64.5	3.0 %	2.1 %	1.5 %	2.5 %	1.8 %	1.6 %
	75 % > income ≥ 50 %	5.6 %	5.4 %	64.3	4.3 %	2.6 %	2.8 %	3.4 %	4.2 %	2.8 %
	50 % > income ≥ 25 %	15.3 %	13.4 %	64.2	3.8 %	2.6 %	3.6 %	3.2 %	4.0 %	3.8 %
	25 % > income	29.8 %	26.9 %	64.4	3.6 %	2.6 %	3.1 %	3.6 %	2.6 %	2.8 %
WOMEN	Income ≥ 75 %	4.1 %	4.6 %	64.4	4.9 %	2.8 %	2.4 %	5.8 %	1.7 %	3.3 %
	75 % > income ≥ 50 %	10.5 %	11.1 %	64.1	7.4 %	2.2 %	3.2 %	7.0 %	1.3 %	2.9 %
	50 % > income ≥ 25 %	18.5 %	20.7 %	64.0	8.3 %	1.5 %	3.3 %	6.9 %	1.6 %	3.5 %
	25 % > income	11.4 %	13.8 %	64.0	8.7 %	1.4 %	2.3 %	7.3 %	1.6 %	4.0 %
		100.0 %	100.0 %							

Table A4. Characterization of persons who retire at 63 in the treatment group, as compared to all other persons in the treatment group. Note: All variables are measured one year before retirement.

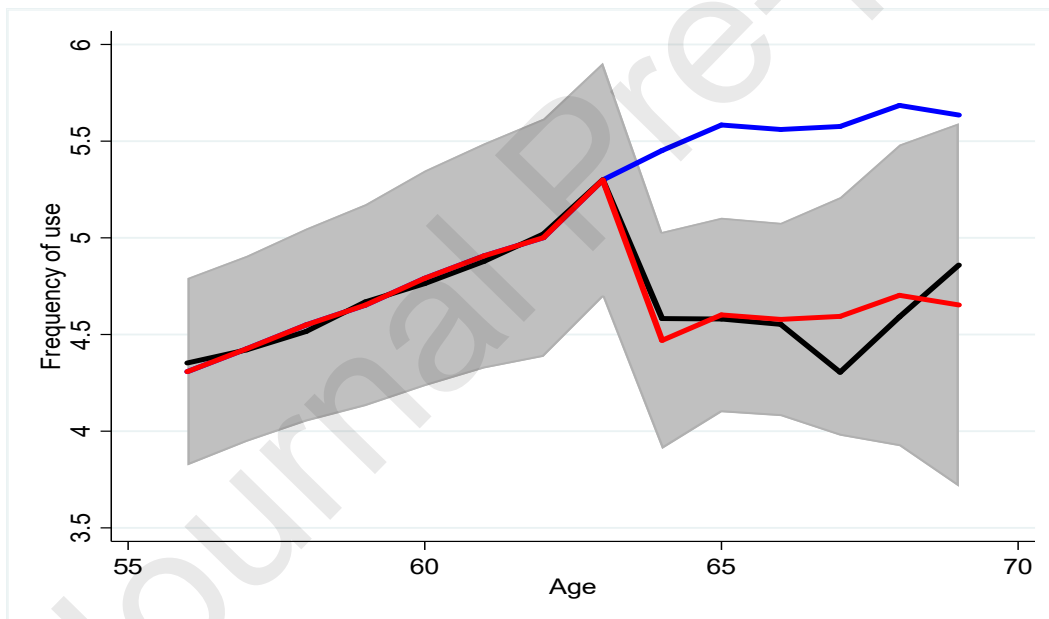## The effects of aging at the aggregate level

Figure 4. Aggregate frequency of antidepressant use (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63. Note: Robust standard errors for heterogeneity and person-level clusters in the error terms. In the results, y-axis denotes the % risk of using antidepressants or facing a hospital period. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.
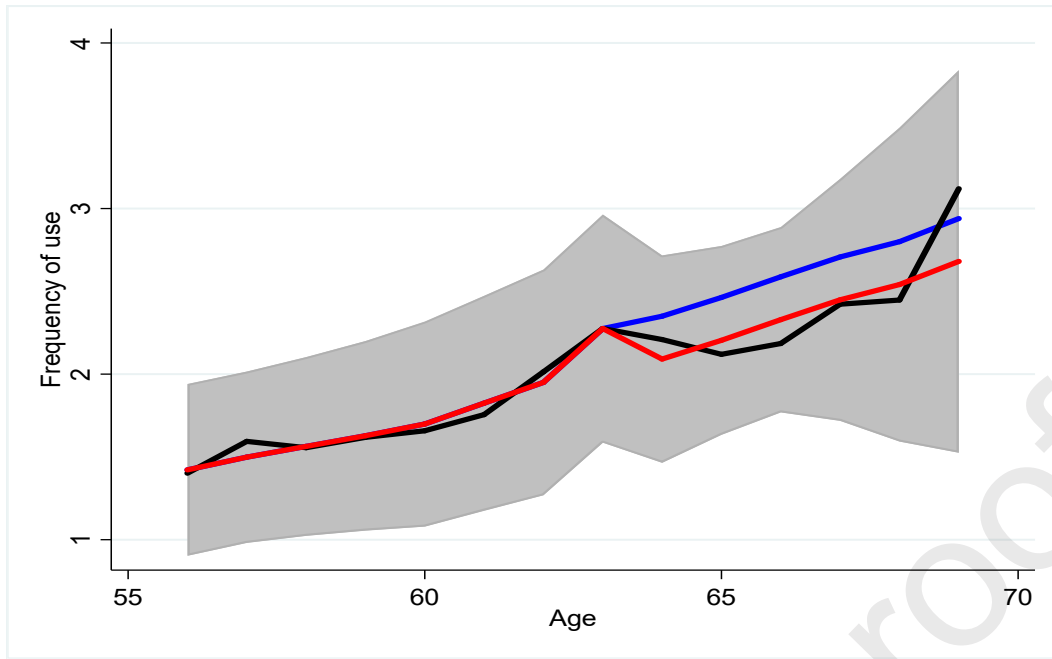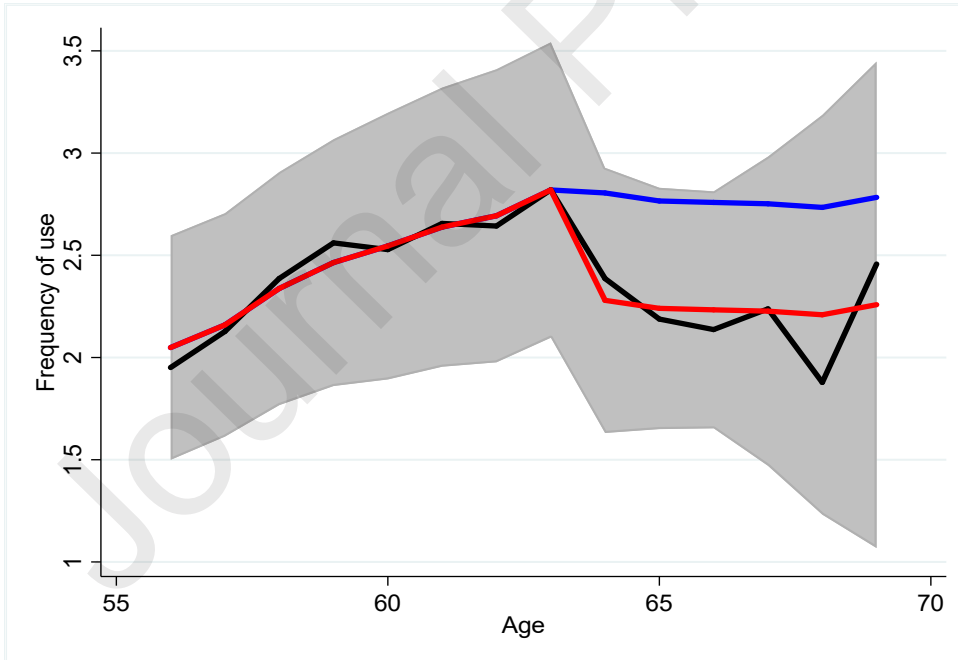
Figure 5. Aggregate frequency of cardiovascular treatment episodes (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63. Note: Robust standard errors for heterogeneity and person-level clusters in the error terms. In the results, y-axis denotes the % risk of using antidepressants or facing a hospital period. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.

Figure 6. Aggregate frequency of musculoskeletal treatment episodes (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63. Note: Robust standard errors for heterogeneity and person-level clusters in the error terms. In the results, y-axis denotes the % risk of using antidepressants or facing a hospital period. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.

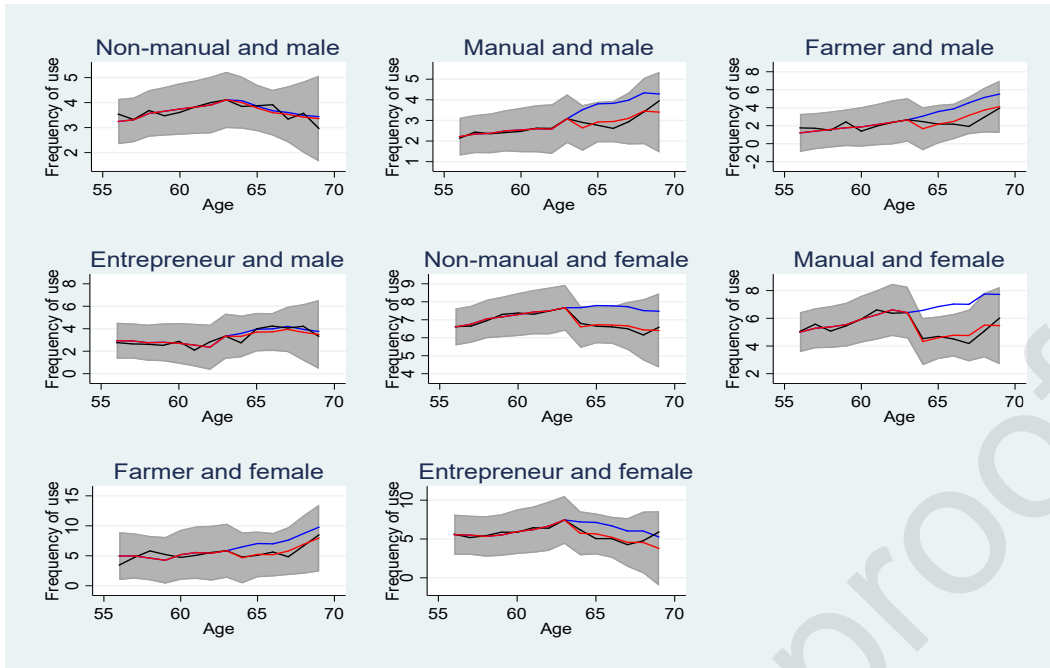## The effects of aging for individual groups

Figure 7. Aggregate frequency of antidepressant use by socioeconomic group (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63. Note: Robust standard errors for heterogeneity and person-level clusters in the error terms. In the results, y-axis denotes the % risk of using antidepressants or facing a hospital period. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.
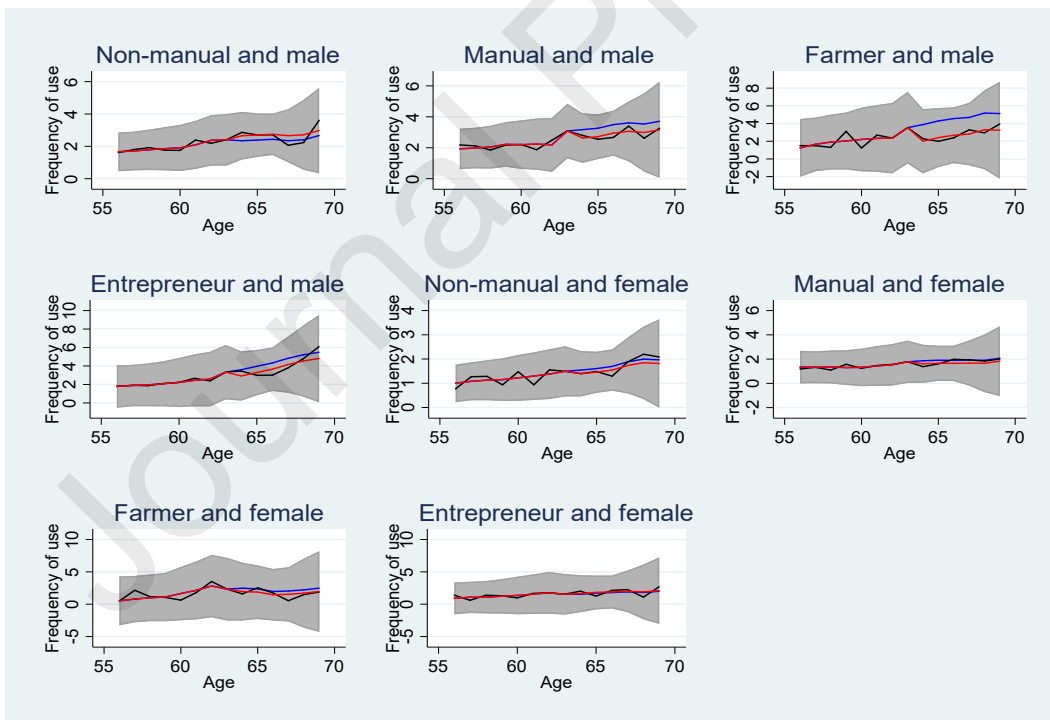

Figure 8. Aggregate frequency of cardiovascular treatment episodes (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profileI without the effect of turning 63. Note: Robust standard errors for heterogeneity and person-level clusters in the error terms. In the results, y-axis denotes the % risk of using antidepressants or facing a hospital period. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.
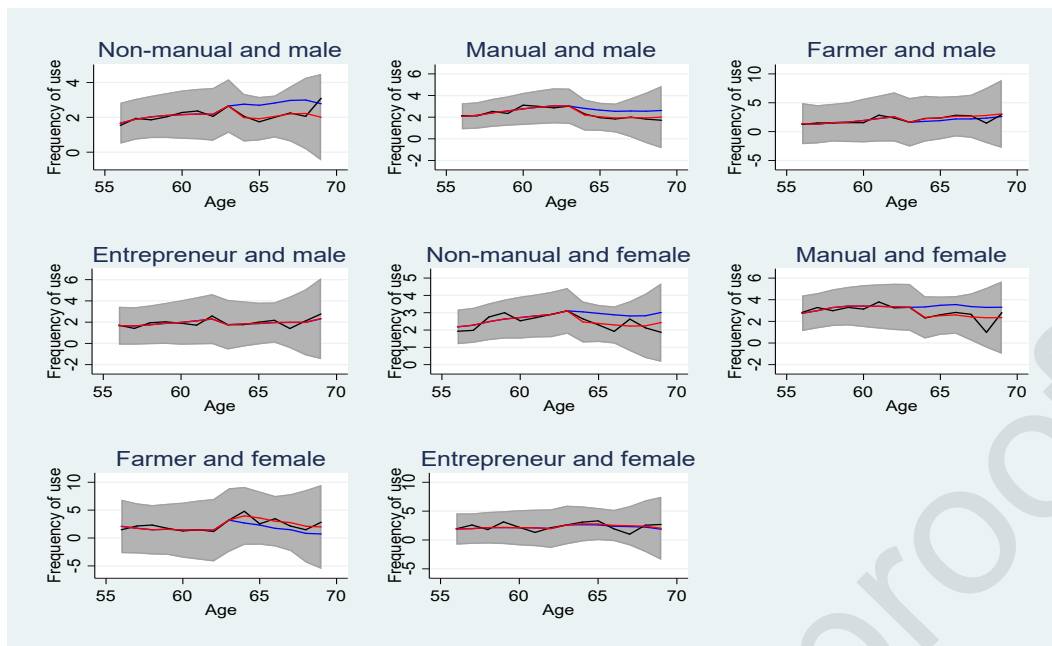
31

Figure 9. Frequency of musculoskeletal treatment episodes by socioeconomic group (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63. Note: Robust standard errors for heterogeneity and person-level clusters in the error terms. In the results, y-axis denotes the % risk of using antidepressants or facing a hospital period. Other control variables: year- and individual-specific fixed effects and step-wise quadratic aging trend.

References

Angrist, J., Imbens, G. W., and Rubin. D. (1996). Identification of Causal Effects Using Instrumental Variables, Journal of the American Statistical Association, 91(434), 444-472.

Angrist, J.D., and Pischke, J.-S. (2009). Mostly Harmless Econometrics: An Empiricists Companion. Princeton, NJ: Princeton University Press.

Angrist, J.D., and Pischke, J.-S. (2015). Mastering metrics: The path from cause to effect. Princeton University Press, Princeton, NJ.

Atalay, K., and Barrett, G. (2014). The causal effect of retirement on health: New evidence from Australian pension reform. Economics Letters, 125(3), 392–395.

Behncke, S. (2012). Does retirement trigger ill health? Health Economics, 21, 282–82.

Belloni, M., Meschi, E., and Pasini, G. (2016). The effect on mental health of retiring during the economic crisis. Health Economics, 25, 126–26.

Bíró, A. (2016). Outpatient visits after retirement in Europe and the US. Int J Health Econ Manag (2016) 16:363–385.

Bíró, A. and Elek, P. (2018). How does retirement affect healthcare expenditures? Evidence from a change in the retirement age. Health Economics, 27(5), 803-818.

Bloemen, H., Hochguertel, S. and Zweerink, J. (2017). The causal effect of retirement on mortality: Evidence from targeted incentives to retire early. Health Economics, 26(12), 204-218,

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

Brinch, C.N., Mogstad, M., Wiswall, M. (2017). Beyond LATE with a discrete instrument. Journal of Political Economy, 125(4), 985-1039.

Charles, K.K. (2004). Is retirement depressing? Labor force inactivity and psychological well-being in later life. Research in Labor Economics, 23, 269–299.

Che, Yi and Li, Xin (2018). Retirement and health: Evidence from China. China Economic Review, 49, 84-95.

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

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

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

Fitzpatrick, M.D., and Moore, T. (2018). The mortality effects of retirement: Evidence from social security eligibility at age 62. Journal of Public Economics, 157, Pages 121-137.

Fonseca, R., Kapteyn, A., and Zamarro, G. (2017). Retirement and Cognitive Functioning: International Evidence. In: Financial decision making and retirement security in an aging world. Edited by Olivia, S., Mitchell, P., Hammond, B., and Utkus, S.P. Oxford University Press.

Gorry, A., Gorry, D. and Nataraj Slavov, S. (2018). Does retirement improve health and life satisfaction? Health Economics, 27(12), 2067-2086.

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

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, D., Johansson, P. and 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,

Hämäläinen, J., Isometsä, E., Sihvo, S., Kiviruusu, O., Pirkola, S., Lönnqvist, J. (2009). Treatment of major depressive disorder in the Finnish general population. Depress Anxiety. 2009;26:1049–1059.

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

Hernaes, E., Markussen, S., Piggott, J., and Vestad, O.L. (2013). Does retirement age impact mortality? Journal of Health Economics, 32(3), 586–98.

Hessel, P. (2016). Does retirement (really) lead to worse health among European men and women across all educational levels? Social Science & Medicine, 151(C), 19–26.

Imbens, G. W., and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. Journal of Economic Literature, 47 (1): 5-86.

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

Kajitani, S., Sakata, K., and McKenzie, C. (2016). Occupation, retirement and cognitive functioning. Ageing and Society, 37(8), 1568-1596

Kleibergen, F., and Paap, R., (2006). Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics, 133, 97–126.

Kolodziej, I.W.K. and García-Gómez, P. (2019). Saved by retirement: Beyond the mean effect on mental health. Social Science & Medicine, 225, 85-97.

Laaksonen, M., Metsä-Simola, M.N., Martikainen, P., Pietiläinen, O., Rahkonen, O., Gould, R., Partonen, T. and Lahelma, E. (2012). Trajectories of mental health before and after old-age and disability retirement: A register-based study on purchases of psychotropic drugs. Scandinavian Journal of Work and Environmental Health, 38(5):409-417.

Lee, D.S., and Lemieux, T. (2010). Regression discontinuity designs in economics. Journal of Economic Literature, 48, June 2010, 281–355

Leinonen, T., Lahelma, E., and Martikainen, P. (2013). Trajectories of antidepressant medication before and after retirement: The contribution of socio-demographic factors. European Journal of Epidemiology, 28(2), 417.

Lemieux, T., and Milligan, K. (2008). Incentive effects of social assistance: A regression discontinuity approach. Journal of Economics, 142(2), 807.

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

Maimaris, W., Hogan, H., and Lock, K. (2010). The impact of working beyond traditional retirement ages on mental health: Implications for public health and welfare policy. Public Health Review, 32, 532.

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

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

Meng, A., Nex, M.A., and Borg, V. (2017). The impact of retirement on age related cognitive decline – A systematic review. BMC Geriatrics, 21(17), 160.

Messe, P-J. and Wolff, F-C. (2019). Healthier when retiring earlier? Evidence from France. Applied Economics, 51:47, 5122-5143.

Mokyr Horner, E., and Cullen, M. (2016). The impact of retirement on health: Quasi-experimental methods using administrative data. BMC Health Service Research, 16(68).

Nielsen, N.F. (2019). Sick of retirement. J Health Econ. 65:133-152.

Nishimura, Y., Oikawa, M., and Motegi, H. (2018). What explains the difference in the effect of retirement on health? Evidence from global aging data. Journal of Economic Surveys, 32(3), 792-847

Oksanen, T., Vahtera, J., Westerlund, H., Pentti, J., Sjösten, N., Virtanen, M., Kawachi, I., and Kivimäki, M. (2011). Is retirement beneficial for mental health? Antidepressant use before and after retirement. Epidemiology, 22, 553 –39.

Oshio, T., and Kan, M. (2017). The dynamic impact of retirement on health: Evidence from a nationwide ten-year panel survey in Japan. Preventive Medicine, 100, 287–293.

Petterson-Lidbom, P. (2012). Does the size of the legislature affect the size of government? Evidence from two natural experiments. Journal of Public Economics, 96(3), 269–269.

Rohwedder, S., and Willis, R.J. (2010). Mental retirement. Journal of Economic Perspectives, 24(1), 119–138.

Schaffer, M.E. (2010). xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. Statistical Software Components S456501, Boston College Department of Economics.

Stock, J.H., and Yogo, M. (2005). "Testing for weak instruments in linear IV regression." In: Identification and inference for econometric models: Essays in honor of Thomas Rothenberg (pp. 80–108). Edited by Andrews, D.W.K., and Stock, J.H. Stock. Cambridge University Press.

Tumino, A. (2016). Retirement and cognitive abilities. ISER Working Paper Series, No. 2016-06. Institute for Social and Economic Research.

van der Heide, I., van Rijn, R.M., Robroek, S.W.J., Burdorf, A., and Proper, K. (2013). Is retirement good for your health? A systematic review of longitudinal studies. BMC Public Health, 13, 1180.

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

Zhu, R. (2016). Retirement and its consequences for women's health in Australia. Social Science & Medicine, 163(C), 117–125.