ETLA Working Papers | 67 18.4.2019

The Influence of Old-age Retirement on Health:

CAUSAL EVIDENCE FROM THE FINNISH REGISTER DATA

Tero Kuusi
The Research Institute of the Finnish Economy
tero.kuusi@etla.fi

Pekka Martikainen
University of Helsinki
pekka.martikainen@helsinki.fi

Tarmo Valkonen
The Research Institute of the Finnish Economy
tarmo.valkonen@etla.fi

Suggested citation:
Kuusi, Tero, Martikainen, Pekka & Valkonen, Tarmo (18.4.2019). "The Influence of Old-age Retirement on Health: Causal Evidence from the Finnish Register Data".

ETLA Working Papers No 67.
http://pub.etla.fi/ETLA-Working-Papers-67.pdf

Abstract

We quantify the impact of old-age retirement on health using longitudinal Finnish register data for the period 2000–2012, which 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 moderately decreases the use of antidepressants, especially for women; (2) the beneficial effects of retirement on the cardiovascular and musculoskeletal conditions are smaller and more diffused; (3) there is no robust evidence that retirement effects vary systematically among socioeconomic groups, although more robust declines in musculoskeletal diseases were observed among manual-labour men; and (4) the beneficial effects in antidepressant use can be extended to apply to most Finns retiring at ages 62–64 based on our test of external validity.

Tiivistelmä

Vanhuuseläkkeelle siirtymisen vaikutuksia terveyteen: Näkökulmia suomalaisesta rekisteriaineistosta

Tutkimuksessa selvitetään työstä vanhuuseläkkeelle jäämisen vaikutuksia henkiseen ja fyysiseen terveyteen käyttäen suomalaisia yksilöaineistoja vuosilta 2000–2012. Kattavan rekisteriaineiston avulla voidaan seurata yksityisen sektorin työntekijöiden siirtymiä suoraan työstä eläkkeelle. Tälle ryhmälle alin täyden vanhuuseläkkeen saamisen ikä, eli 63 vuotta, on merkittävä eläköitymisen taloudellinen kannuste. Syy-seuraussuhteen varmistamiseksi käytetäänkin FE-IV -estimointimenetelmää, jossa instrumenttimuuttujana on kyseinen 63 vuoden ikäraja.

Tulosten mukaan vanhuseläkkeelle siirtyminen vähentää masennuslääkkeiden käyttöä jonkin verran erityisesti naisilla. Tuki- ja liikuntaelinsairauksien ja sydän- ja verisuonitautien osalta positiiviset vaikutukset ovat vähäisemmät ja hajanaisemmat. Sosioekonomisten ryhmien välillä ei havaita järjestelmällisiä eroja, vaikka tuki- ja liikuntaelinsairauksien väheneminen onkin selvintä työntekijämiehillä. Lisäanalyysit tulosten yleistettävyydestä viittaavat siihen, että masennuslääkkeiden käytön vähenemistä koskeva tulos voidaan yleistää 62–64-vuotiaina eläkkeelle siirtyneille.

We gratefully acknowledge research funding from the Academy of Finland (decision No. 303649).

Kiitämme Suomen Akatemiaa taloudellisesta tuesta (apurahan numero 303649).

Keywords: Health, Health behaviour, Retirement, Retirement policies, Demographic economics

Asiasanat: Terveys, Terveyskäyttäytyminen, Eläkkeelle siirtyminen, Eläkepolitiikka, Väestö, Talous

JEL: I10, I12, J26, J18

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. It is partially unclear whether thew3 declines in mortality are matched with corresponding increases in health and functioning. 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 still remains in the subject.

In this paper, our research agenda is two-fold. First, we provide new local average treatment effect (LATE) estimates for the health effects of retirement using the individual Finnish population, retirement and health registers. Second, we test and extend the external validity of these results to a broader age group of retirees. The importance for the generalisation of the results arises because LATE typically identifies the treatment effects for only a minority of the retirees, which has weakened the policy relevance and generalisability of the previous results.

Our individual register data allow us to identify clear transitions from employment to retirement, where all the elements of the transition (not working, having more leisure, having less income) are present. The framework is strengthened by the fact that the earnings-related first pillar pension scheme, which we analyse, dictates in practise the retirement income and retirement decisions in Finland. A total of 93% 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.

The data included here comprise information on all prescription medications and contacts with the hospital care system of the studied individuals (N = 93,381) in 2000–2012. Therefore, we avoid the problems related to using self-assessed health measures. 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. A further advantage of the data is that non-response or loss to follow-up biases do not affect them.

We address the health–retirement endogeneity problem using the instrumental variable (IV) approach. The lowest statutory eligibility age for full old-age pensions in Finland, 63 years, was a good predictor of the true retirement age during the studied period. Moreover, 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. We also consider individual time-invariant fixed effects.

We find that retirement decreases the use of antidepressants, whereas the effect on physical conditions is smaller and more diverse. While the aggregate impacts are relatively small, we report significant heterogeneity across sociodemographic groups: a moderate decline in antidepressant use among the women and a decline in hospital episodes due to musculoskeletal diseases among the manual workers. The results for cardiovascular diseases are similar, but they are not statistically significant in most groups.

We also use novel methodology to address the generality of our results. One of the key challenges of current retirement health literature is that IV methodology only provides the LATE at exactly the 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. The increase in the probability of retirement due to reaching age 63 is at best around 30%. Meanwhile, the majority of retirements from work (circa 68%) occur between the ages of 62–64. Thus, being able to address the heterogeneity of the treatment effect in this region considerably increases the generalisability of our results.

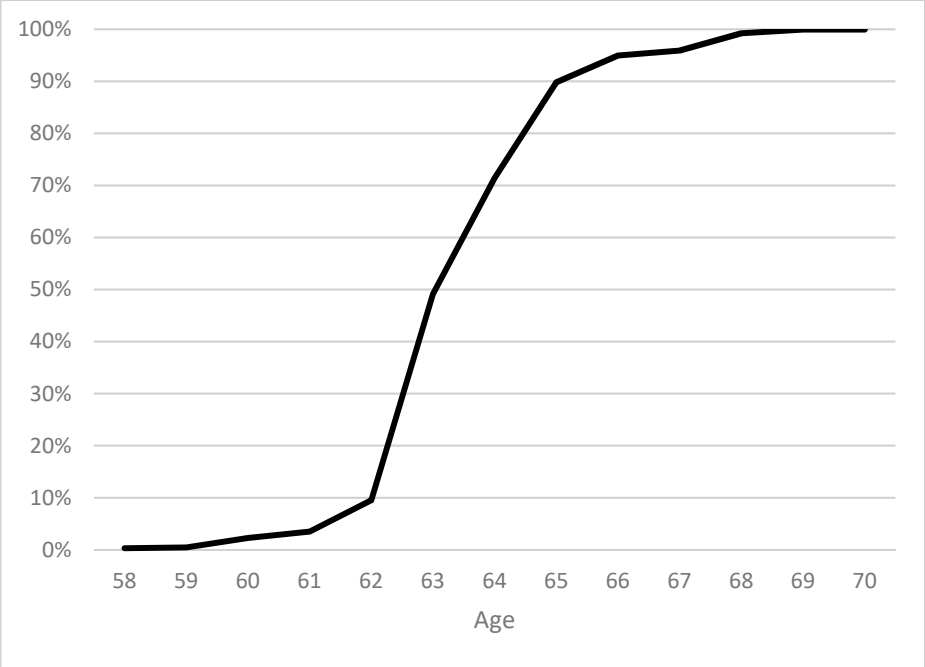

Figure 1. Cumulative distribution of retirement by age for individuals who retire from work to an old-age pension in the private sector from 2000–2012 in Finland.

We test the heterogeneity of the marginal effects of retirement on our health variables in the different segments of the population. In particular, we construct two test statistics: (1) We compare the health outcomes at the year of retirement of those who retire at ages 62 and 63 (treated test statistics). (2) We compare the health outcomes at ages 62 and 63 of those who will retire at 63 and 64, correspondingly (untreated test statistics). If the two test statistics are equal, it is a sign of treatment homogeneity.

We find evidence against the homogeneous treatment hypothesis. Our general finding is that the homogeneity assumption can be rejected when comparing health outcomes far from the threshold and if the time-invariant individual-level heterogeneity is not properly accounted for. Instead, at the age window of 62–64 years (when most Finns retire), 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 generalise our LATE results concerning these health outcomes to the broader age window.

Having said this, we find some evidence suggesting that the effect of retirement on the cardiovascular hospital treatments is heterogeneous, even at the close age interval of 62–64 years. We find that people aged 62 who retire at 62 are healthier than people aged 63 who retire at 63 when compared to their long-term average health. In contrast, people who retire at age 63 are typically as healthy at age 62 as people aged 63 who retire at 64. These findings suggest that there is retirement selection at age 63, and that the selection effects are heterogeneous in a manner that may result in upward bias in the effect of retirement.

Finally, we address other typical problems of the model design. We estimate the short-term effect at the proximity of the retirement age and the longer-term average effect. We test the 'placebo' influence of the age at the proximity of the statutory retirement age for people who do not retire. Furthermore, we address selection by resorting to matching. We conclude that our results are not sensitive to these potential problems.

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.[1]

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. 2017) found that the choice of statistical methods and cofounders has a strong influence on the results.

The latter type of studies shows that retirement either weakens (Rohwedder and Willis 2010, Mazzonna and Perachhi 2012, Bonsang et al. 2012, Tumino 2016, Nishimura et al. 2017) or has 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 weakens retirement effects.

Results regarding mental health show that the impact of retirement is either positive (Atalay and Barrett 2014, Zhu 2016, 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.

The main result from the previous studies analysing the physical effects of retirement is that health improves, but there is heterogeneity in the findings. Insler (2014) shows that health improves 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. Mazzonna and Peracchi (2017) find that health improves in physically demanding occupations, but in occupations where physical stress is low, retirement may weaken health. Hessel (2016) reports heterogeneity when analysing European data – retirement is beneficial to 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.

On the one hand, Nielsen (2018) finds no effect of statutory retirement age on physical health, but the popular early retirement seems to reduce general practitioner (GP) episodes and hospitalisation in Denmark. On the other hand, a study from Norway (Hernaes et al. 2013) shows that early retirement has no impact on mortality. Similar conclusions are drawn in a Swedish study analysing an increase in the statutory retirement age (Hagen 2018). Negative effects have also been detected. Behncke (2012) shows that retirement increases chronic conditions and weakens self-assessed health in the UK. Fitzpatrick and Moore (2017) notice that retirement increases male mortality in the USA.

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

[1] 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.

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 these results can arise from many sources. The designs and the statistical methods used may be stronger or weaker in their ability to identify the causal relationships between retirement and health. Measures of health vary from self-assessed general indicators to register-based detailed diagnoses and from use of health services to mortality. 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. The problems are aggravated if the data that are used are collected from several countries with different types of pension schemes and other welfare systems, such as SHARE. As is common in statistical studies, cofounders also vary.

In this article, we address several potential caveats in identifying effects that may result in differences between our and the earlier results.

Our analysis suggests that one should be very careful when interpreting the results for studies that use the statutory retirement age as an instrument. That is because in most of the existing literature, reaching the statutory retirement age leads to only partial selection of retirement (of treatment). Some people choose to retire before the retirement age, and others who reach it forgo the possibility of retiring. Individuals who reach the earliest eligibility age for full old-age pension are in the intervention group. They receive an intervention that affects selection into treatment. Individuals who have not reached this age are in the control group, and they do not receive the intervention. However, all individuals can choose to receive or to forgo the treatment. Only a fraction of the intervention group, the so-called 'compliers', take the treatment, while the rest of the intervention group, the 'never takers', do not. Meanwhile, the so-called 'always takers' take the treatment even if they have not yet reached the statutory full pension age. Thus, the research design is quasi-experimental, with two-sided noncompliance (Brinch et al. 2017)

Noncompliance is problematic when the effect of the intervention is not strong, but rather there is only a small increase in the probability of entering retirement due to reaching the statutory age (i.e. the share of compliers is low). While the IV estimation provides an unbiased estimate of the LATE of the compliers, 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 compliers is only a fraction of the population, it is a particular group that may be especially prone to either the positive or negative health effects of retirement.

Furthermore, we aim to avoid the 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 discontinuities used as instruments (Lee and Lemieux 2010), and we study the sensitivity of the findings by narrowing the range of the data analysed around the discontinuities (Angrist and Pischke 2015) in order to sufficiently explore the non-linear effects of 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.

Both the definition of retirement and the statutory retirement age have often been unclear. 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 any clear incentives for retirement can induce additional measurement errors. 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

intervention group would improve the first stage results of IV, but it would distort the estimates of the health effects of retiring from work.

Instead, we can isolate the pure retirement transition from work to old-age pension with our register data. This group is the only one that is relevant 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. The case of Finland represents a clean institutional case since the earnings-related first pillar pension scheme dominates retirement income and retirement decisions – the second occupational pillar and individual third pillar with their different retirement rules are marginal.

In previous studies, measures of health vary from self-assessed general indicators to register-based detailed diagnoses and from use of health services to mortality. Even though self-assessed general health may be a rather comprehensive indicator, its accuracy is questionable, especially when multi-country data are used. We use health register data, which includes all the studied individuals' prescriptions and contacts with public health care. More precisely, we measure the influence of retirement on mental health by purchases of antidepressants; we measure the impact of retirement on physical health by number of hospital episodes due to cardiovascular or musculoskeletal diseases.

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. 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.

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 and who are residents of Finland. Furthermore, 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; 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. 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.

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 I00finedand musculoskeletal diseases as M00 mus.

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 of 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 occupational and personal retirement ages remained the same 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. 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. 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 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).

Formally, 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:

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

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

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

and

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

where $\overline{sp}_{it}$ is an indicator variable that receives a value of 1 if $age_i \geq sp_i$ and receives a value of 0 when $age_i < sp_i$. 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 ($\beta_3, \beta_4, \beta_5, \alpha_3, \alpha_4, \alpha_5 = 0$).

In the estimations, $\hat{r}_{it-1}$ is the prediction of r_{it-1} from the first stage, with $\overline{sp}_{it-1}$ used as an excluded instrument. Thus, our baseline model estimates the average effect of retirement to the health variable at the year that follows the retirement and after that.[2] We use the year- and person-fixed effects, ϵ_t and ϵ_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.

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).

In this respect, the Finnish pension system provides good guidance: In the first stage, we use the age limit for the eligibility for full old-age pension benefits as an IV. In the second stage, we explain health by the expected propensity to retire at age 63 estimated in the first stage. Formally, the instrument is the indicator variable

[2] 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.

$turned63_{it} = I(t \geq the\ year\ when\ person\ i\ turns\ 63)$. We chose the instrument 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 at age 63).

β_1 represents the difference in the probability of treatment between the control group and the intervention group. The difference is often called the first stage; it gives an estimate of the impact of the retirement on the fraction of treated individuals. Typically, the research generates a LATE – that is, $\frac{\alpha_1}{\beta_1}$. 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).[3]

However, even when the standard tests reject weakness of the instrument, the external validity of the LATE may be violated when $\beta_1$is small and there is evidence against the homogeneity of the expected marginal treatment effects in the population. In this case, the group of compliers is selected and the selection matters for the outcome. 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 instrument of turning the eligibility age). This would invalidate the generality of the LATE outside the complier group.

To measure the heterogeneity, we investigate the generality of our results within the age interval of 62–64. Generalisations of the results is relevant because the majority (68%) of the Finnish private sector employees retire from work at this age. Accordingly, the first test statistically 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 never takers (have turned 63 but will retire at 64). Because it compares the untreated outcomes, these test statistics can be referred to as the untreated outcome test statistics. If the statistic is different from zero, the mean outcomes of the untreated compliers are different from the never takers. 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 always takers (aged 62 and retired in the same year) and compliers (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*. Those who have preferences toward taking the treatment (early) are different from the compliers.

[3] 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.

In our framework, we define the untreated and treated test statistics, Δ_0 and Δ_1, respectively, as follows:

$$\Delta_0 = E[health_{it}|age_{it} = 63\,, retires\ at\ 64] - E[health_{it}|age_{it} = 62\,, retires\ at\ 63]$$
$$= \Delta_0 age + \Delta_0 E[\epsilon_i] + \Delta_0 E[\epsilon_{it}] \quad (5)$$
$$\Delta_1 = E[health_{it}|age_{it} = 63\,, retires\ at\ 63] - E[health_{it}|age_{it} = 62\,, retires\ at\ 62]$$
$$= \Delta_1 age + \Delta_1 E[\epsilon_i] + \Delta_1 E[\epsilon_{it}] \quad (6)$$

The test statistics capture several effects: (1) the mean effect of aging, Δage, which we assume to be common to all groups; (2) the mean time-invariant differences in the health outcomes between the groups, $\Delta E[\epsilon_i]$; and (3) the mean differences in the model's idiosyncratic errors.

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.[4] 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:

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

where the first term is expected to be zero by assumption, 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 amount of compliers and decreases the amount 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.

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.

[4] 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.

Furthermore, we have access to a long time series of personal health outcomes that allows us to properly control for person-level fixed effects.

In addition, our rich dataset allows us to control other possible problems within the statistical inference. We avoid 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.

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 is at best low. Thus, it appears that retirement does not affect the health status of the treatment group relative to the health status of the same-aged control group for the same year.

VARIABLES	Use of antidepressants	Hospital periods (cardiovascular)	Hospital periods (musculoskeletal)
Retirement	**-0.00413*** **	-0.000436	3.60e-05
	(0.00127)	(0.000838)	(0.000835)
Age	-	-	-
Age^2	-1.81e–06	**4.83e–05*** **	**1.41e–05** **
	(1.22e–05)	(6.80e–06)	(6.68e–06)
Observations	1,148,465	1,148,465	1,148,465
Number of individuals	93,381	93,381	93,381

*Table 2. Fixed-effects panel estimation of the aggregate effect of retirement on health. The coefficient of the retirement variable denotes the effect of retirement on health. In the results, -0.01 denotes a 1% decrease in the risk of using antidepressants or facing a hospital period. 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 industry-specific fixed effects. We omitted the linear age trend due to collinearity.*

We then repeat the estimation using the IV methodology in Equations (1) and (2). The results in Table 3 suggests 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.[5]

[5] 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.

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 the effect of turning 63 is roughly a 1% decline in the risk of using antidepressants and a 0.5% decline in the risk of a hospital period due to musculoskeletal disease. We further illustrate the predictions of the reduced-form model and their uncertainty in Figures 1–3 in the Appendix. We show that the decline in the reduced-form estimates fits well with the actual changes in health outcomes, and the model provides reasonable counterfactual dynamics for health outcomes in the absence of turning 63.

VARIABLES	**Use of antidepressants**	**Hospital periods (cardiovascular)**	**Hospital periods (musculoskeletal)**
Second stage			
LATE, α_1/β_1	**-0.0305*** **	**-0.008***	**-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
N (obs)	210,465	210,465	210,465
N (people)	17,635	17,635	17,635

*Table 3. IV estimation of the aggregate effect of retirement on health. In the results, -0.01 denotes a 1% 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 industry-specific fixed effects. We omitted linear age trend due to collinearity.*

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. 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.

We first study the socioeconomic groups. In terms of antidepressant use, the first finding is that the effect of retirement is to systematically decrease the likelihood of antidepressant use. For some groups, the effect is stronger than for others – but in all cases, the sign is negative, indicating a decrease in antidepressant use after retirement. We have also analysed the effects 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.

VARIABLES		Second stage: Use of antidepressants		Second stage: Hospital periods (cardiovascular)		Second stage: Hospital periods (musculoskeletal)		First stage				
		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)	N (people)
MEN	Non-manual	-0.0024	0.0103	0.0098	0.0115	**-0.0244****	0.0107	**0.3167*****	0.0084	1405	4409,1	3,700
	Manual	**-0.0222*****	0.008	-0.0138	0.0105	-0.0155	0.0105	**0.3968*****	0.0093	1815	38,214	3,212
	Farmer	**-0.0731***	0.0375	**-0.0983***	0.0503	0.0244	0.0433	**0.1912*****	0.0196	95	8,080	678
	(Other) entrepreneur	-0.0074	0.0126	-0.019	0.0177	0.0004	0.0138	**0.3545*****	0.0134	698	18,921	1,588
WOMEN	Non-manual	**-0.0362*****	0.0119	-0.0051	0.0083	**-0.0197***	0.0112	**0.2952*****	0.0072	1690	60,099	5,023
	Manual	**-0.0685*****	0.0148	-0.0075	0.011	**-0.0292***	0.016	**0.3285*****	0.0104	989	26,866	2,246
	Farmer	-0.0887	0.0544	-0.0248	0.0475	0.0623	0.0648	**0.2042*****	0.0264	60	4,069	342
	(Other) entrepreneur	-0.0472*	0.0244	0.0031	0.0198	0.005	0.0259	**0.3108*****	0.0182	291	10,125	846
MEN	Single	-0.019	0.0168	0.0203	0.021	-0.0223	0.0179	**0.3094*****	0.014	486	16,640	2,144
	Non-single	**-0.013****	0.0061	**-0.0156****	0.0077	**-0.0137***	0.0073	**0.345*****	0.006	3320	92,666	8,357
WOMEN	Single	**-0.0725*****	0.0208	-0.0147	0.013	-0.0165	0.0178	**0.2591*****	0.0099	684	28,873	3,215
	Non-single	**-0.0383*****	0.0092	-0.0034	0.0072	**-0.0188***	0.0101	**0.3157*****	0.0067	2191	72,286	6,882
MEN	Income ≥ 75%	-0.0002	0.0189	**-0.0481***	0.0258	-0.0257	0.0183	**0.3205*****	0.0196	267	9,111	769
	75% > income ≥ 50%	-0.0105	0.019	-0.0357	0.0245	-0.0182	0.0222	**0.3166*****	0.0171	341	11,383	958
	50% > income ≥ 25%	**-0.034*****	0.0106	0.0055	0.0132	-0.0019	0.0127	**0.3614*****	0.0105	1187	29,508	2,478
	25% > income	-0.0044	0.0078	-0.0086	0.0096	**-0.0211****	0.0091	**0.3459*****	0.0075	2138	58,709	4,919
WOMEN	Income ≥ 75%	-0.038	0.0304	0.0088	0.0236	0.0393	0.028	**0.2734*****	0.0183	224	9,220	771
	75% > income ≥ 50%	**-0.053*****	0.0168	-0.0027	0.0112	-0.0002	0.0146	**0.3469*****	0.0113	938	22,793	1,907
	50% > income ≥ 25%	**-0.0498*****	0.0127	-0.0102	0.0098	**-0.0368*****	0.0135	**0.3089*****	0.0086	1296	41,740	3,487
	25% > income	**-0.0461****	0.0193	-0.0045	0.013	-0.0189	0.0199	**0.2649*****	0.0111	568	27,031	2,259

*Table 4. The effect of retirement on health by group. 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 industry-specific fixed effects. We omitted linear age trend due to collinearity.*

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. Table 1

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. Finally, the risk of facing a hospital period due to a musculoskeletal disease lowers especially in the lower-income quantiles.

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. For single women, the effect is substantial. However, the generalisability of this result is brought into question by the possibility of selection (see Table 6).

We next test the heterogeneity of the treatment effects. We first measure the individual test statistics without controlling for the individual-level fixed effects according to Equations (5) and (6). In all cases, 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:

$$\Delta_0 = E[health_{it}|age_{it} = 63\,, retires\ at\ age \geq 64] - E[health_{it}|age_{it} = 62\,, retires\ at\ 63]$$
$$= \Delta_0 age + \Delta_0 E[\epsilon_i] + \Delta_0 E[\epsilon_{it}] \quad (7)$$

$$\Delta_1 = E[health_{it}|age_{it} = 63\,, retires\ at\ 63] - E[health_{it}|age_{it} = 62\,, retires\ at\ age \leq 62]$$
$$= \Delta_1 age + \Delta_1 E[\epsilon_i] + \Delta_1 E[\epsilon_{it}] \quad (8)$$

We again reject homogeneity for all health outcomes.

Instead, when we control for the individual-level fixed effects, 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.

VARIABLES	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
Δ_1 - Δ_0	0.0039	**0.0088***	0.0068
T-test statistic Δ_1 - Δ_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.

Let us consider this result in more detail. The rejection of the homogeneous treatment hypothesis in the case of cardiovascular hospital treatments is rejected 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. Instead, when we allow for comparison beyond the age interval 62-64, that is, we use Equations (7) and (8), the homogeneity of the treatment effects is rejected at the 95% confidence interval for all health variables.

All in all, the results indicate that the use of individual fixed effects to control for heterogeneity may be important, and that once this has been done, the heterogeneity is of a moderate size.

We also repeat the analysis for sub-groups of individuals. We find that the joint hypothesis of heterogeneity across the different groups (chi-squared distributed F-test statistics with degrees of freedom equal to the number of sub-groups) can only be rejected for the cardiovascular hospital periods by socioeconomic class. While the test may suffer from weak power in the sub-groups, it is worth mentioning that the significant rejections are typically observed for groups that have insignificant results also in terms of p-values.

VARIABLE		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.

6. Robustness analysis

Finally, we consider robustness checks in addition to our main results (Table 7). First, one possible caveat is that selection to retirement, and thus our LATE estimate, may be reflecting short-term changes in health. To control for them, we include to our estimations the measures of health outcomes one year prior to the retirement as additional control variables in the specification ‘‘Previous year's health outcomes as additional controls’’. In order to analyze the robustness of our results in terms of timing, we 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 the specification ‘‘Contemporaneous effect of retirement’’. Furthermore, to ensure that our results are not due to the stratification of the data, 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 order to 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’’. We also use coarsed exact matching to preprocess our data in order to avoid the potentially confounding influence of pre-treatment control variables, and thus reducing imbalance between the treated and control groups in the specification ‘‘Estimation after matching’’. Accordingly, we first 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, and repeat the analysis. Finally, we repeat the analysis in a small window 62–64 around the statutory retirement age while controlling for the age effects with a linear time trend in the specification for the age effects window 62–64.

All in all, our robustness analysis is in line with our main results. 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.

Specification	Second stage						First 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	β_1	SE	F -test	N (obs)	N (people)
Previous year's health outcomes as additional controls	**-0.0267*** **	0.0048	**-0.008***	0.0048	**-0.0161*** **	0.0053	**0.1694*** **	0.0052	6879	210,465	17,635
Contemporaneous effect of retirement	**-0.0354*** **	0.0076	**-0.0113***	0.0062	**-0.0195*** **	0.0064	**0.2146*** **	0.0051	5554	228,126	17,635
All retirements included	**-0.0346*** **	0.0031	-0.0048	0.003	**-0.0163*** **	0.003	**0.0375*** **	0.0022	26204	1,054,824	92,731
The aggregate effect on men	**-0.0146** **	0.0057	-0.0099	0.0071	**-0.0149** **	0.0066	**0.2024*** **	0.0074	3887	109,306	9,178
The aggregate effect on women	**-0.0497*** **	0.0087	-0.0058	0.0062	**-0.0178** **	0.0086	**0.1329*** **	0.0074	3003	101,159	8,457
Estimation after matching	**-0.0394*** **	0.0113	-0.0144	0.013	-0.0226	0.0139	**0.3124*** **	0.004	951	188,682	15,799
Estimation at the age window 62–64	**-0.0261*** **	0.0073	**-0.0195***	0.01	-0.0096	0.0088	**0.2869*** **	0.0045	4037	51,053	17,632

*Table 7. Robustness analysis. IV estimation of the aggregate effect of retirement on health. In the results, -0.01 denotes a 1% 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.*

7. Conclusions

Our objective in this study was to determine whether retirement from work to old-age pension influences various health indicators, and whether this impact differs between 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 beneficial effects of retirement on the cardiovascular and musculoskeletal conditions are smaller and more diffused. Overall, we do not find robust evidence that these retirement effects vary systematically among socioeconomic groups, although we observe more robust declines in musculoskeletal diseases among manual 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 at ages 62–64.

Many of the previous studies, including some that use the Finnish data, did not detect an effect of 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. Our results confirm the importance of using the data and identification strategy in analysis. The observed causal connection between physical health and retirement among men in manual occupations is in line with the results of the few previous studies that consider physical requirements of different jobs.

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. On the one hand, we found that encouraging retirement may not be the most efficient policy to reduce the common health conditions studied among 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.

Appendix

The effects of aging at the aggregate level

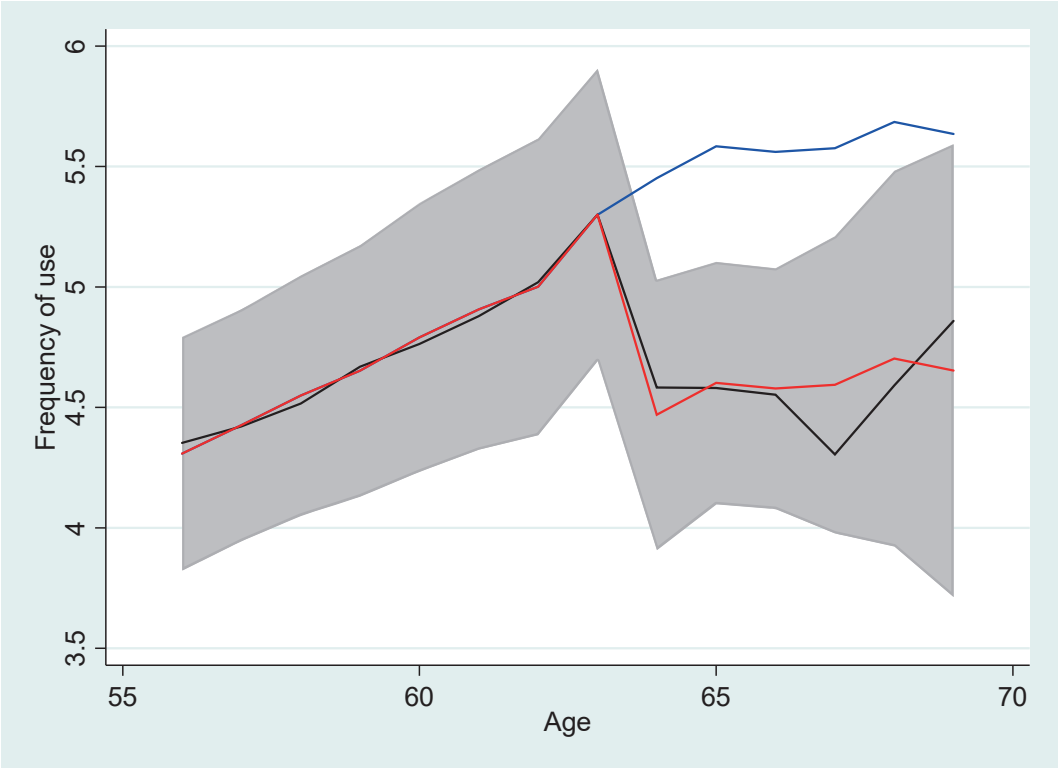

Figure 1. Aggregate frequency of antidepressant use (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63.

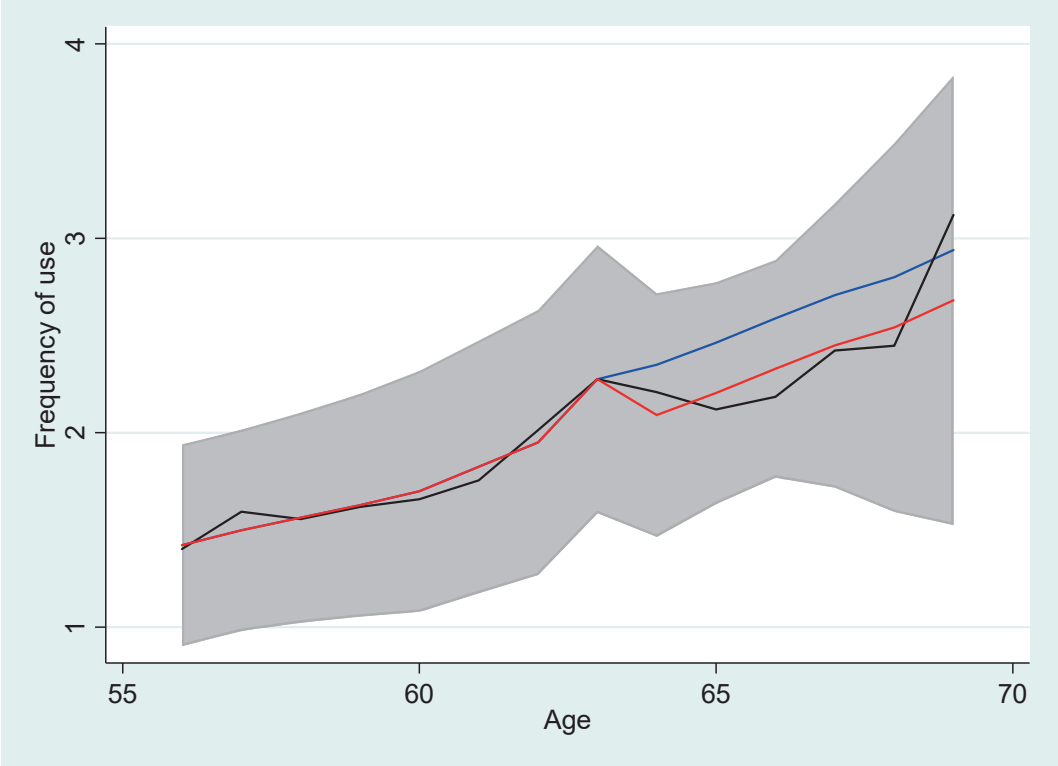

Figure 2. Aggregate frequency of cardiovascular treatment episodes (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63.

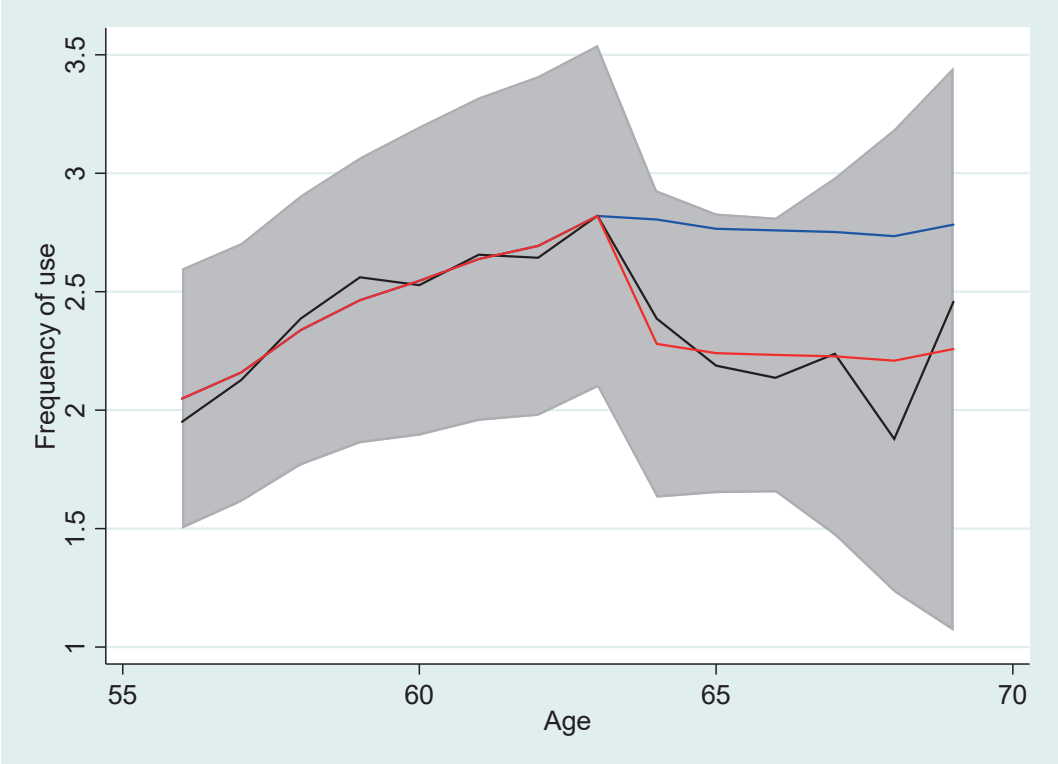

Figure 3. Aggregate frequency of musculoskeletal treatment episodes (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63.

The effects of aging for individual groups

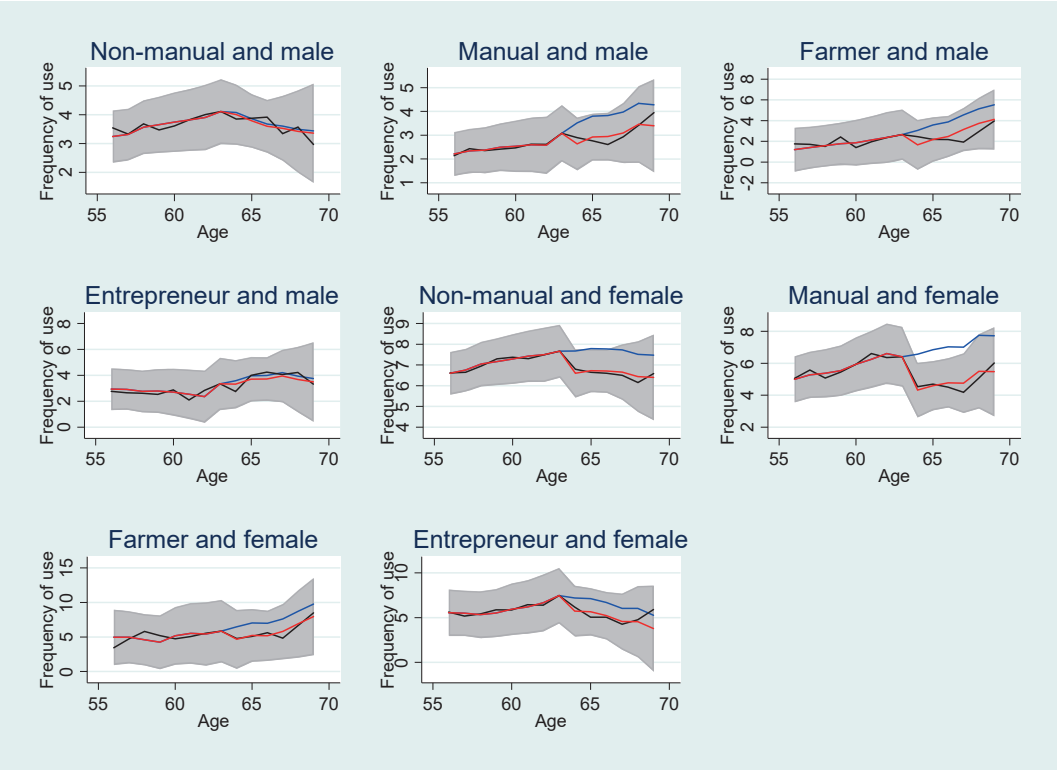

Figure 4. Aggregate frequency of antidepressant use by socioeconomic group (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63.

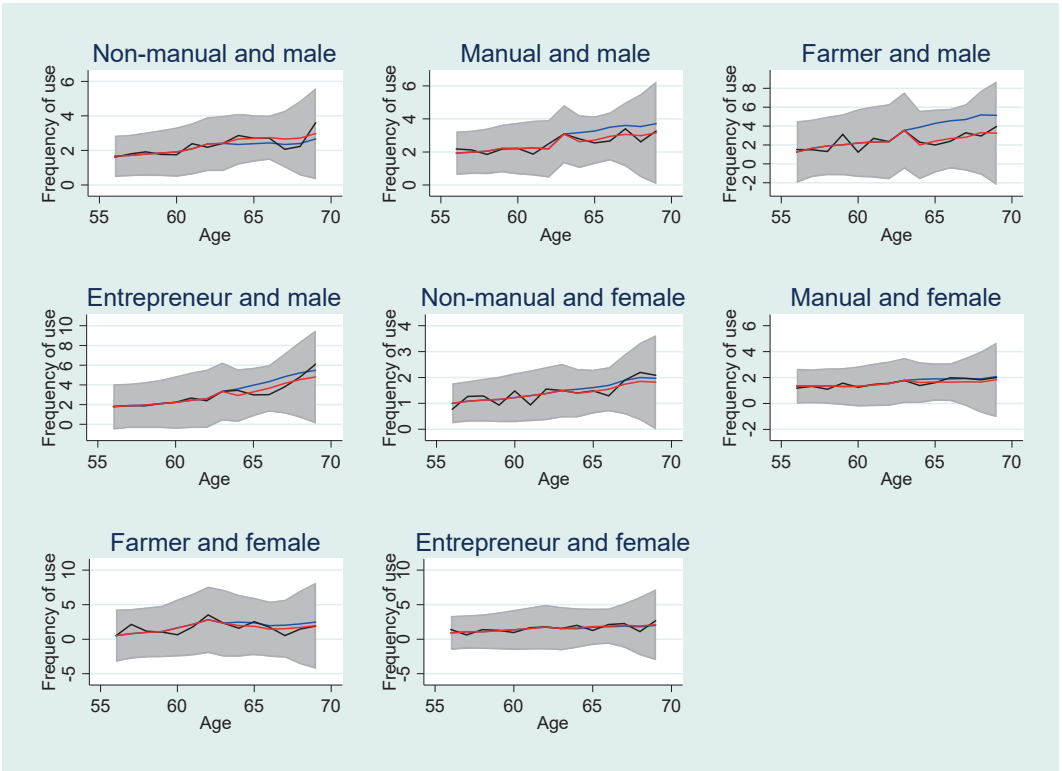

Figure 5. Aggregate frequency of cardiovascular treatment episodes (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profilel without the effect of turning 63.

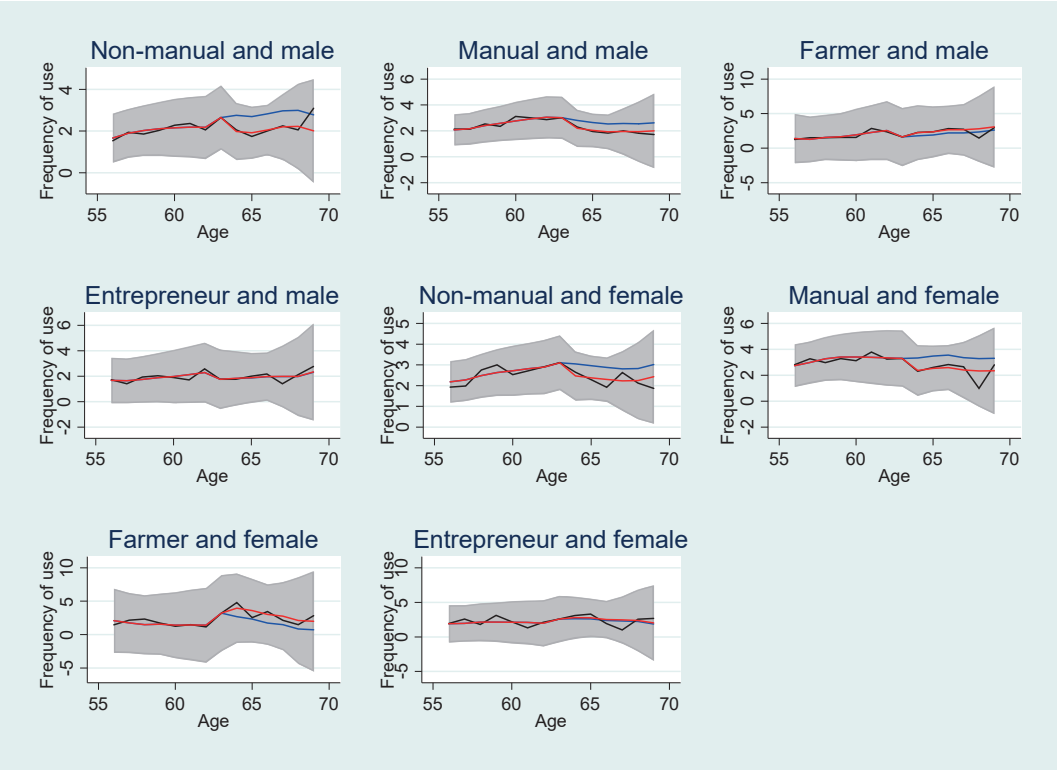

Figure 6. Frequency of musculoskeletal treatment episodes by socioeconomic group (black line), prediction of the reduced-form model (red line) and its 95% confidence interval (grey area). The blue line denotes the counterfactual age profile without the effect of turning 63.

References

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.

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.

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.

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

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. (2017). The mortality effects of retirement: Evidence from social security eligibility at age 62. NBER Working Paper No. w24127.

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.

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.

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.

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.

Kloep, M and Henry, KB. (2006). Pathways into retirement: Entry or exit? Journal of Occupational and Organizational Psychology, 79, 569–593.

Laaksonen, M., Mets, M.N., Martikainen, P., Pietilietil, 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.

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.

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. (2018). Sick of retirement. Unpublished manuscript Available at https://ssrn.com/abstract=3096512

Nishimura, Y., Oikawa, M., and Motegi, H. (2017). 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., Sjosten, N., Virtanen, M., Kawachi, I., and Kivimmi, 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.

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

Elinkeinoelämän tutkimuslaitos

The Research Institute of the Finnish Economy

ISSN-L 2323-2420
ISSN 2323-2420 (print)
ISSN 2323-2439 (pdf)

Tel. +358-9-609 900
www.etla.fi
firstname.lastname@etla.fi

Arkadiankatu 23 B
FIN-00100 Helsinki