The causal effect of retirement on health

WORKING DRAFT, DO NOT CITE OR DISTRIBUTE

Nick Fabrin Nielsen*

April 10, 2017

Abstract

This paper examines the causal effect of retirement on health. Through extensive information from Danish administrative records, I am able to track the health and socioeconomic status of all Danish residents born before 1960. To estimate the causal effect of retirement on health, two approaches are taken. First, I use a reform-induced change in the retirement eligibility age only affecting individuals born on, or after, July 1st 1939. This allows me to use individuals born before the threshold as a control group. Second, I use a large discontinuity in retirement takeup at age 60 - causing an increase in retirement takeup of 17 percentage points - in a regression-discontinuity design. For both approaches, there is no significant effect of retirement on the Charlson index or mortality but the RD design shows that healthcare utilization decreases as a result of retirement.

Keywords: Retirement, health, Instrumental Variables (IV) Estimation
JEL classification: I10, C26, J26

*Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, DK-1353 Copenhagen, and Centre for Applied Microeconometrics (CAM), Denmark. Email: nick.fabrin.nielsen@econ.ku.dk, Tel: +45 35337574. I thank Mette Ejrnæs, Carl-Johan Dalgaard, Josh Goldstein, Will Dow and seminar participants at UC Berkeley, University of Copenhagen and QED Jamboree Amsterdam for helpful discussion and comments

1 Introduction and Motivation

The world population is aging. In the next 40 years, the total share of the population aged 60 or older is expected to rise from around 10% of the global population to 22% of the global population (Bloom et al., 2015). In Denmark, this ratio is expected to reach 30% by 2053 (Statistics Denmark, 2016). For governments and policymakers alike, population aging is a cause for concern due to the inherent social and economic challenges it brings. To alleviate the fiscal burden of aging populations, some policies seek to increase employment amongst the oldest, for example by incentivizing delayed retirement or changing statutory retirement ages (Whiteford and Whitehouse, 2006). These policy changes have direct effects on government budgets through increases in tax payments and decreases in government transfers. However, retirement alterations may also influence the health of individuals and therefore impact fiscal spending. If retirement has large effects on health, these effects, both fiscal and in terms of individual well-being, should be taken into account in the evaluation of retirement reforms. This paper sheds light on this important issue by examining the causal effect of retirement on health and healthcare utilization for Danes born before 1960.

In the workhorse model of Grossman (1972), retirement may have both positive and negative impacts on health. In Grossman's model, retirement affects the health stock via its potential impact on health investments. Retirement may change individual health investments through two channels: 1) it alters the productivity incentives of health investments and 2) it changes the marginal value of time. In terms of changes in productivity incentives, retirement lowers the incentive to invest in health by eliminating the income effects from sick days on the labor market. Post-retirement, there is no direct adverse income effects from being ill as there is no sickness-pay or reduced wages due to illness in retirement. Furthermore, once an individual has retired, there is no indirect effects running through decreased promotion opportunities or lower productivity due to fewer days of job experience. The reduction in the productivity based incentives to invest in health will, ceteris paribus, lead to declining levels of health after retirement. The other potential effect on health investments stems from post-retirement changes in the marginal value of time. Changes in the marginal value of time has ambiguous effects on health, regardless of the direction of change. As an example, consider a decrease in the marginal value of time due to the increase in leisure time availability after retirement. The decrease will simultaneously reduce the time cost of investing in health (e.g. by exercising, cooking healthy food or visiting the GP more regularly for health examinations) and reduce the value of healthy leisure time (making it less costly to be ill in terms of foregone leisure time). The reduction in the costs of health investments will lead to rising health levels, whereas the reduction in the marginal value of healthy leisure time will have the opposite effect. Within the Grossman model, we cannot determine which of these countervailing effects will dominate. This holds true even if we are willing to assume a direction of the change in the marginal value of time - as exemplified above. Naturally, retirement may also affect health through channels which are not adequately captured in the Grossman model. For instance, poor working conditions may have adverse effects on health which would be affected by retirement. Another potential channel could be social networks, where the loss of daily contact with collegues after retirement could affect mental health or well-being. Ultimately, the lack of a clear theoretical prediction leaves the question of the health effects of retirement open for empirical investigation.

A growing body of empirical research has investigated the impact of retirement on health. However, in spite of the growing attention, there is no consensus on the sign or the size of the effect of retirement on health because of mixed results which tend to vary according to differences in empirical methods or operationalization of health. Of the papers finding non-negative effects of retirement on health, several exploit political reforms or adjustments for identification (Hallberg, Johansson and Josephson, 2015; Atalay and Barrett, 2014; Bloemen, Hochguertel and Zweerink, 2017; Bingley and Pedersen, 2011) while others use age discontinuities in regression discontinuity designs (Eibich, 2015; Johnston and Lee, 2009) or use statutory retirement ages as instruments for retirement (Coe and Zamarro, 2011). Some of the papers finding negative effects of retirement on health also apply statutory or early retirement ages as instruments for retirement (Godard, 2016; Heller-Sahlgren, 2017), while others rely on the conditional independence assumption (Dave, Rashad and Spasojevic, 2008) or a mix of these (Behncke, 2012). In a recent survey article, Motegi, Nishimura and Oikawa (2016) investigate the main reasons for the varying findings in the literature on the effects of retirement on health, pointing towards differences in econometric strategy as the main driver behind the discordant results. Although differences in econometric design does seem to play an important role with respect to the empirical differences, the choice of health measures may also be quite important. As an example, Atalay and Barrett (2014) use an Australian retirement reform affecting old age pension (OAP) eligibility ages for Australian women to instrument retirement. They show that retirement has significant positive effects on (i.e. reduce the prevalence of) back pain and mood disorders but no significant effects on prevalence of high stress, heart conditions or hypertension. It is not clear whether the absence of statistical significance for some measures of health is due to imprecise estimates or if retirement only affects specific parts of health, but the example shows that choice of health measure has empirical importance. Ultimately, the lack of agreement across studies, even with rather similar strategies or measures of health, calls for further explorations into the effects of retirement on health.

This paper applies two distinct strategies to identify the causal effect of retirement on the health of Danes: a retirement reform and a large discontinuity in retirement takeup. Because I use the same third-party reported measures of health in both empirical designs, I can assess how results vary with respect to choice of strategy and therefore help bridge the gaps in the previous findings.

I investigate two direct or "pure" measures of health and two measures of health capturing a broader mix of health status and healthcare utilization. The pure measures of health are the Charlson comorbidity index and 1-year mortality while the health utilization variables are hospitalizations and the number of visits to the GP. The Charlson index and mortality are broad measures of overall health and are therefore rather similar to the health stock in the Grossman model[1]. The hospitalization and GP visit variables capture healthcare utilization more broadly, enabling investigations into the effects on both health status and health investments. Furthermore, knowledge of the effects on healthcare utilization is relevant for evaluations

[1]Other broad measures of health can be found in the medical literature, for example the frailty index in Mitnitski et al. (2002). The Charlson index has previously been applied in the economics literature, making comparisons across papers easier.

of the fiscal impact of retirement.

The first strategy exploits an exogenous change in the Danish Old Age Pension (OAP) age following a retirement reform enacted in 1999. One part of the reform was a change of the minimum OAP age from age 67 to 65. The change only affected individuals born on or after July 1st 1939, providing an opportunity to use individuals born just before the threshold as a natural control group. Hereafter, I will refer to the strategy as the reform strategy. The second identification strategy utilizes a large jump in retirement takeup at age 60, caused by the Danish Early Retirement Pension, to indentify the causal effect of retirement on health in a regression-discontinuity-design (RDD). I will refer to this strategy as the RD strategy.

The empirical results show that both the reform and the discontinuity induce considerable variation in retirement takeup. In the reform strategy, the reform causes a gap in retirement takeup at age 66 of just above 15 percentage points between affected and non-affected individuals. In the RD strategy, individuals at age 60 are more than 17 percentage points more likely to be retired than individuals at age 59. Moving onto the effect of retirement on health, the reform strategy shows no significant effects on any of the measures of health. In the RD strategy, there is no significant effect on the Charlson index or 1-year mortality, but small negative effects on healthcare utilization. The magnitudes of the effects on healthcare utilization are small but highly significant: retirement leads to a drop in the number of GP visits corresponding to 7% of the mean number of GP visits and a drop in the risk of being hospitalized corresponding to 8% of the mean hospitalization rate. The results show that if policymakers *delay* retirement, it will lead to *increases* in public healthcare spending. For public budgets, however, these effects will be small relative to the increases in tax revenue following delays in the retirement age.

This paper proceeds as follows. In the following two sections, I lay out the institutional settings, available data, and sample selection. In the fourth section, I present the reform strategy and its results. In the fifth section, I present the regression discontinuity strategy and results. The sixth section discusses the empirical results and the fiscal importance of the healthcare channel.

2 Institutions

Pathways to retirement In Denmark, there are three main pathways to retirement: Old Age Pension (OAP), Early Retirement Pension (ERP) and Disability Pension (DIP)[2]. These programmes are largely state-funded and pay a monthly benefit to eligible Danish citizens. In addition to these, individuals can save for their retirement through labor market pensions and private pensions.

OAP is available for all Danish citizens aged 65 (67) and older[3]. Changes to the OAP scheme was part of a major retirement reform enacted in 1999. One element of the reform was a change in the OAP age from age 67 to 65. The reform was implemented such that individuals born on or after July 1st 1939 would be able to retire to OAP at age 65, while the OAP retirement age remained at 67 for individuals born before July 1st 1939. Consequently, individuals eligible for ERP, DIP or other

[2]Outside of the main retirement schemes there are smaller programmes such as the Civil Servant Pension. However, these schemes are utilized to a subtantially smaller degree and will not be covered in detail here.

[3]The level of OAP payouts are subject to requirements on the number of years with residence in Denmark

types of pensions born before July 1st 1939 were still able to retire through these channels up until the age of 67. The 1999 reform entailed several changes related to retirement and pension savings, but the change in OAP eligiblity age was the only change which had differential impact according to the birth date threshold.

ERP is available from age 60 until the age of OAP entitlement (65/67) for individuals who have been members of an unemployment insurance fund for a sufficient number of years (10-25 years)[4]. Since its introduction in 1979, ERP has been quite popular. In 2003, for example, 52% of the 64 year olds were on ERP.

DIP is available for individuals in the working age with permanently reduced work capacity and is granted by municipalities on basis of an assessment of work capacity, health and social circumstances. In addition to the formal retirement paths, individuals can choose informal retirement through the unemployment system or by leaving the labor force altogether.

Healthcare in Denmark The Danish health care system is, for all important purposes, state funded with universal coverage[5]. For Danish residents, visiting a general practitioner, going to a specialist or being admitted to hospital is not subject to any out-of-pocket expenses, insurance coverage requirements, maximum limits or similar preconditions. Main exceptions are dental care and medicine purchases

[4]Requirements have changed over time: Until 1992, the requirement was 10 years of membership, from 1992 to 1999 it was 20 years of membership. In 1999 it was changed to 25 years of membership with the addition of a new monthly ERP contribution.

[5]There is private health insurance (typically through the employer), which covers only very specific types of treatments. Treatment through private health insurance is typically minor surgery such as meniscal tear surgery. These treatments are still available through the public system but, compared to private health care, individuals may have to wait longer for public treatment.

which require out-of-pocket spending (but are subsidized[6]).

3 Data

I have access to several administrative registers, which enables me to construct a dataset with comprehensive information on health and socioeconomic status for all Danish residents born before 1960 for the years 1980-2010. Individuals cannot self-select into or out of these datasets, so there is only attrition due to emigration and death.

For information on health related issues, I have access to in-patient hospital records[7], and the central death registry. Furthermore, I have access to payment records for general practioners and specialists from 2005 onwards. The hospital records contain information on diagnoses, hospital admittance and duration recorded for research and funding purposes by medical professionals such as nurses, medical doctors and medical secretaries. Patients have no discretion over these registrations.

I use diagnosis information for diagnoses labeled as either main diagnoses or secondary diagnoses (there can be one or more of both types). Diagnoses are only recorded if they are clinically relevant for the specific contact with the hospital. This means that an eye disease, such as cataract, is not registered when a patient is being treated for pneumonia (unless the cataract is deemed relevant for the treatment of

[6]Without special conditions, dental care is subsidized but individual costs can reach high levels. Medicine is subsidized, with subsidy rates increasing with the total yearly medicine costs. Currently, the highest yearly out-of-pocket medicine expense is roughly 4,000 DKK (approximately 600 USD).

[7]From 1995 onwards, I have access to out-patient records including registrations from emergency rooms. The registration of out-patients has changed substantially over time, therefore I disregard this information in the analysis.

pneumonia by the medical professionals). In such as case, however, the cataract will be registered upon treatment of the cataract. Diagnoses are registered according to the International Classification of Diseases (ICD) version 8 until 1993 and version 10 from 1993-2010.

From the GP payment records, I cannot see specific diagnoses but only the number of procedures performed by doctors and specialists. Each procedure has a unique identifying number which can be used to calculate the number of visits to the GP. Any single consultation can consist of several procedures that will require registration. For example, a visit to the GP could consist of three procedures: the consultation itself, renewing a prescription and taking blood samples. To count the number of actual visits to the GP, I only count the consultation procedures (excluding email or phone consultations). The appendix contains the specific codes used.

For information on income, retirement and education I link several Danish administrative datasets. The majority of information is based on tax records. As with the health data, information in these registers is based on reports from third parties such as tax authorities, financial institutions, employers or educational institutions. The socioeconomic data is generally measured with great precision.

Health and healthcare utilization As dependent variables, I use the Charlson comorbidity index, a dummy for hospitalization, the number of GP visits and 1-year mortality.[8]

[8] Because I have access to detailed hospital diagnosis codes, I could technically investigate the effects of retirement on several hundred measures of health (even thousands). There are two connected reasons not to pursue this idea: 1) missing theoretical predictions and 2) multiple hypothesis testing issues. First, the lack of a model that has different predictions of the effects of retirement on specific measures of health, makes results very difficult to interpret. This is especially important in combination with a high risk of false discoveries. Second, a multiple hypothesis testing issue arises because

From the central death registry, I can observe mortality directly measured on a daily basis. Because my retirement variable is registered for survivors only, I construct a measure of mortality which is compatible with the retirement variable. Specifically, I construct an indicator capturing if individuals die in the following calender year,

$$\text{Mortality}_{i\,t+1} = \mathbf{1}(\text{Jan. 1st, year } t+1 \leq \text{death}_i \geq \text{Dec. 31st, year } t+1) \qquad (1)$$

This measure will capture the effects on mortality in the calendar year following retirement but will not allow me to see any effects of retirement on mortality in the year of retirement.

I use hospital records to measure hospitalizations and to calculate the value of the Charlson index. For hospitalizations, I set a hospitalization dummy equal to "1" if individual i has been admitted to a hospital in year t and "0" otherwise.

The Charlson comorbidity index weighs different serious illnesses by the degree of their severity. For instance, an individual with Diabetes type II (weight 1) and a metastatic solid tumor (weight 6) has a Charlson co-morbidity index value of 7. I calculate a yearly value of the Charlson comorbidity index based on the ICD adaptation by Johansen and Fynbo (2011), i.e.

$$\text{Charlson}_{it} = \sum_{d=1}^{\bar{d}} \omega_d \mathbf{1}(\text{diag}_{it} = d) \qquad (2)$$

of the probability of false positives when several hypotheses are tested without proper adjustment of the standard errors. Naturally, there are techniques for multiple hypothesis testing adjustments (see e.g. Shaffer (1995)) but these are often not without problems of their own. Therefore, I focus on the Charlson index and mortality as overall measures of health and hospitalization and GP visits as variables capturing healthcare utilization.

where d indexes all Charlson diagnoses[9] for individual i in year t. The appendix contains the specific ICD-8 and ICD-10 codes used to construct the index. The Charlson index or dummies derived from the Charlson index has previously been applied to measure health in the economics literature by Contoyannis et al. (2005) and Nielsen (2016).

I use the payment records to count the number of visits to the GP for each individual in each year as my measure of GP utilization.

Control variables I base my definition of retirement on the Danish Register-Based Labor Force Statistic (RLS). RLS is an administrative measure, dividing the entire Danish population into employment categories following the guidelines of the International Labour Organisation (ILO) adapted for use with administrative registers. For RLS, employment status is tallied every year in the last week of November on basis of a battery of different registers on the Danish population. Employment status is not measured for deceased individuals. I follow Statistics Denmark and define retirees as individuals who are receiving ERP, OAP, DP or civil servant pension. Information on income and education is collected from the tax and education registers.

For sample selection purposes in the OAP reform strategy, it is necessary to determine whether individuals are eligible for ERP once they reach age 60. This variable is unobserved but can be constructed with reasonable precision. Up un-

[9]The Charlson comorbidity disease categories are (adapted from Johansen and Fynbo (2011)): $\omega_d = 1$: Myocardial infarction, Congestive heart failure, Peripheral vascular disease, Cerebrovascular disease, Dementia, Chronic pulmonary disease, Connective tissue disease, Ulcer disease, Mild liver disease, Diabetes type 1 and 2. $\omega_d = 2$: Hemiplegia, Moderate to severe renal disease, Diabetes with end organ damage type 1 and 2, Any tumor, Leukemia, Lymphoma. $\omega_d = 3$: Moderate to severe liver disease. $\omega_d = 6$: Metastatic solid tumor, AIDS.

til 1999, ERP eligibility were solely determined from the total number of years the individual had been a member of an unemployment insurance fund. In 1999, an additional monthly ERP contribution was added to the eligiblity requirement. Both unemployment insurance fund membership and contribution payments are observed. Therefore, I can construct a variable measuring whether or not an individual is eligible for ERP at age 60. I define an individual to be eligible for ERP if I can observe at least 10 years of UI fund membership before age 60. The choice of the 10 year horizon is a compromise between sample size and precision of the eligibility measure. For instance, the 1998 ERP eligibility requirement was 20 years of UI fund membership[10]. If I were to choose a 20 year horizon, I would not be able to use observations from 1998 as observations do not start until 1980. To assess how well the eligibility measure works, table A.1 presents a tabulation of ERP eligibility and ERP takeup at age 60, 62 and 64 for the 1939 cohort. False negatives are in the north-east corner of each sub-tabulation. At age 60, 3% of the individuals labeled not-eligible are on ERP. At age 64, the takeup rate for non-eligible individuals were 12% and the take-up rate for eligible individuals 71%. The figures in table A.1 demonstrate that the share of false-negatives is limited, suggesting the usefulness of the ERP eligibility measure.

4 Reform strategy

In 1999, a major retirement reform was enacted. One part of the reform was a change of the minimum OAP age from age 67 to 65. The change only affected

[10]Technically, the 1998 requirement varies somewhat across individuals because ERP was not introduced until 1979 and due to the phase-in of the 1992 reform.

individuals born on or after July 1st 1939. This particular implementation of the reform allows for estimation of causal effects, using differences in exact birth date to instrument retirement. Because the reform in the OAP age affects individuals differently according to a birthday threshold on July 1st 1939, I limit the sample to the 1939 cohort. I exclude individuals I can determine to be eligible for ERP, in which case individuals can retire already at age 60. I do this for two reasons: 1) individuals on ERP are already retired and thus cannot re-retire when they reach the OAP eligiblity age, and 2) individuals eligible for ERP should not be affected by the OAP eligibility age as they have the option to retire to ERP before reaching OAP age (even if they choose not to use this option). With this exclusion, I zoom in on the individuals I expect to be most responsive (with respect to actual retirement behavior) to the policy-induced change in the OAP age. As a further exclusion, I remove individuals who receive DIP before the age of 60 which, for the 1939 cohort, is the age at the enactment of the 1999 reform. I allow for individuals to receive DIP after age 60, as this may be an alternative retirement route for the least healthy non-ERP eligible individuals. I allow for this alternative retirement route to avoid implicitly selecting healthier individuals in the group born before July 1st 1939[11].

Table 1 contains descriptive statistics for the full sample and the sample used in the reform strategy. Due to the aforementioned sample restrictions, individuals in the reform sample are mechanically all 55 years old in 1994, not on DIP at age 55

[11]To see why excluding all individuals receiving DIP could be problematic, consider the decision to apply for disability pension (DIP) at some age after 60. The incentives to apply for DIP are largest for the group born before July 1st 1939, as they can otherwise expect to remain in the labor force until age 67 (instead of 65). Therefore, a larger fraction of individuals born in early 1939 may be expected to apply for, and thus being granted, DIP. By excluding these individuals, I would be excluding a larger share of least healthy from the group with OAP age 67 compared to OAP age 65.

and not eligible for ERP at age 60. The reform sample are more likely to have a college degree, be self-employed, be married, be in the labor force and have higher incomes than the population as a whole[12]. Furthermore, they are healthier measured by the Charlson comorbidity index and mortality and less frequent users of the Danish healthcare system. Table 1 also shows that individual health deteriorates with age as well as healthcare utilization increases with age.

4.1 Reform analysis

I use the difference between individuals born on either side of July 1st 1939 to estimate the causal effect of retirement on health and healthcare utilzation in a straightforward IV design. Formally, the model is,

$$\mathrm{H}_{it} = \beta_0 + \beta_1 \mathrm{R}_{it} + \mathrm{X}_{it}\gamma + u_{it} \tag{3}$$

$$\mathrm{R}_{it} = \delta_0 + \delta_1 \mathrm{Z}_{it} + \mathrm{X}_{it}\alpha + \nu_{it} \tag{4}$$

where H_{it} is a measure of health, R_{it} is a retirement dummy and X_{it} a vector of controls (including age). Equation (4) acts as the first stage with the instrument for retirement $\mathrm{Z}_{it} = \mathbf{1}(birthday \geq 1939_{july1st})$. The identifying assumption is that Z_{it} is uncorrelated with the health shock, u_{it}. Since the reform differs according to season of birth, one may be worried that the identifying assumption is violated due to differences in early life circumstances (see e.g. Moore et al. (1997); Dee and Sievertsen (2015); Buckles and Hungerman (2013)). In the results section, I provide evidence that seasonality is not a concern for the emprical strategy.

[12] A comparison of incomes at age 55 in table 1 across samples is likely to underestimate the difference between the full sample and the reform sample due to inflation

From model (3), $\hat{\beta}_1$ will be the estimate of the average treatment effect for compliers. Compliers are individuals who choose to retire before age 67 because their birthday allows them to retire already at age 65 (but otherwise would have continued to work). Because these individuals are exactly those who adjust their retirement timing in response to changes in the OAP age, their average treatment effect highly relevant for policy evaluation.

4.2 Reform results

Figure 1 shows the retirement ratio by age, split by birthday and is a graphical presentation of the first stage in the reform strategy. It is clear that the likelihood of retiring at age 65 or 66 is much larger for individuals born after July 1st 1939 due to the reform of the OAP age. An important aspect of figure 1, is the strikingly similar retirement patterns between the two groups before age 65 and again after age 67 (when both groups face similar retirement possibilities and incentives). This suggests that it is indeed the OAP reform that drives the retirement difference between the two groups at age 65 and 66. Notice that even though the retirement rates are very similar at age 67 and older, individuals born before July 1st 1939 will, on average, have been retired for fewer years than individuals born after the threshold. This means that if retirement has lasting effects on health (for example by increasing individual health investments) we should be able to see differences in health even after age 67 due to the differences in the number of years in retirement.

Figure 2 shows the average values of the Charlson comorbidity index, the hospitalization rate, the average number of GP visits and the 1-year mortality rate conditional on age and birthday and is a graphical depiction of the reduced form. For all measures of health or healthcare utilization, there are no differences between the two groups before age 65. This is reassuring, as we would not expect to see any differences according to birthday before the reform kicks in. Table B.1 formalizes this, showing mean values and test of equal means or proportions between the two groups at age 55 (before the retirement reform was enacted). The table shows no significant differences between the two groups, again validating the design. At age 65 and beyond, figure 2 shows no visible signs of any difference between the two groups in terms of health or healthcare utilization. This is graphical evidence that there is little or no effects of retirement on health or healthcare utilization for this group of individuals.

Table 2 contains the formal analysis of the effects of retirement on health and healthcare utilzation. The top panel shows the linear regression estimates of health onto retirement without use of instruments, corresponding to regressions of model (3) without using the birthday threshold instrument. For all measures of health or healthcare utilization, we see that retirement is strongly associated with worse health and more frequent use of the healthcare system. Comparing the coefficients in the OLS panel with the mean values, we see that retirement is associated with sharper declines in the direct measures of health (the Charlson index and mortality) than it is associated with increases in hospitalizations and GP visits. This may be due to the fact that hospitalizations and GP visits encompasses more than just individual health (such as health investments). The middle panel shows the reduced form, corresponding to equation (3) inserting the birthday threshold instead of the retirement dummy. This is essentially a formalization of figure 2. As in the graphical analysis, the reduced form regression shows no differences across birth date threshold. The bottom panel shows the IV estimate from model (3) with equation (4) as the first stage. In line with the reduced form estimates, the IV estimates show no effect of retirement on health or healthcare utilization for the reform compliers. From the bottom of table 2 we see the F-statistics from the first stages. With values around 40, we should not worry about weak instrument issues. This which was expected from the large retirement takeup differences visible in the graphical first stage evidence in figure 1.

4.3 Reform robustness

Alternative specifications To check the robustness of the main results, table B.2, B.3, B.4 and B.5 contain estimates from alternative specifications of model (3) with first stage (4) for the Charlson index, a dummy for hospitalization, GP visits and 1-year mortality, respectively. As a baseline, the first columns in each table is a linear regression with a quadratic polynomial in age, complementing the OLS panel in table 2. Columns (2) to (4) contain the IV estimates, adding an increasing number of controls. The estimates in column (4) corresponds to the estimates in the 2SLS panel of table 2. As evident from all tables, there is very little difference in the estimate on retirement across specifications. This points towards the robustness of the estimates presented the main table 2.

As a supplement to the mortality graphs and regressions in figure 2 and table 2, figure B.1 shows the survival rate from age 60 and onwards, split by birthday. Compared to figure 2, figure B.1 contains more detail as it utilizes the high frequency of the Danish central death registry (daily frequency). However, the conclusion does not change. Figure B.1 shows no difference according to birthday before, or after, age 65 and rules out long-term differences in survival according to birth-date.

Placebo cohort To investigate the possible differences between individuals according to season of birth and the potential influence of unobserved covariates, I replicate the analysis using the 1938 cohort instead of the 1939 cohort. For the 1938 cohort there is no difference in OAP eligibility age according to day of birth. Figure B.2 shows the retirement ratio and means of the health variables split by 1938 birthday threshold. The top subfigure shows no difference in the retirement ratio according to time of birth, except for a slightly increased likelihood of retirement at the statutory OAP age (67), where individuals born in the first half of 1938 are more likely to be retired than those born in the second half. This happens because retirement status is measured at the end of each year, leaving more time for the oldest in each cohort to retire before retirement is measured (perhaps individuals retire a few weeks or months after reaching the OAP age to finish projects, train new workers, etc.). Notably, this pattern of slightly increased retirement propensity at age 67 for the oldest in the cohort is reversed compared to the 1939 cohort, where the youngest in the 1939 cohort were more likely to be retired at age 65 than the oldest (because they were eligible for OAP already at age 65). For the 1938 placebo cohort, there is no difference in the average values of health or healthcare utilization across individuals split by birthday threshold. This is reassuring, as it suggests that there is no significant differences in health status according to season of birth, which could otherwise ruin identification. Table B.6 formalizes these placebo graphs by running 2SLS regressions instrumenting retirement with the birthday threshold shifted back to 1938. For all measures of health or healthcare utilization, there is no effect of retirement on health.

5 RD analysis

The Danish retirement rules have very different financial implications for a 59-year-old compared to a 60-year-old. If a 59-year old retires, she will not receive any financial support from the Danish government until she reaches 60, unless she is granted disability pension (which is only available if her work-capacity is permanently reduced). If a 60-year old retires, she can receive ERP if she is eligible, decide to have her pension savings paid out or retire through some of the less popular retirement schemes. Therefore, there is a large discontinuous jump in retirement takeup for Danes at age 60. I use this discontinuity to estimate the causal effect of retirement on health in a regression discontinuity design. This approach has been applied in the literature on retirement and health by e.g. Eibich (2015) (exploiting the German retirement legislation) and Johnston and Lee (2009) (for the UK) - but is also very similar to the cross-country IV method used by Coe and Zamarro (2011). Table 1 contains descriptive statistics for the full sample used in the RD analysis. The full sample contains all Danish residents born before 1960 from 1980-2010, but due to the age restrictions not all of these individuals will be used in the RD regressions and graphs. For instance, I only observe individuals born in 1925-1955 at age 55 and only observe individuals between age 55 and 65 from cohorts 1915-1955. This means that the actual sample used in regressions will be smaller than the total number of observation available at all ages.

5.1 RD strategy

Before I discuss the RDD estimation procedure, I lay out the necessary basic assumptions. First, the dependent variable (health or healthcare utilization) must be a continuous function of the assignment variable (age). I assume that health deteriorates smoothly with age[13] which therefore also applies to the measures capturing physical health directly (Charlson index and mortality). This assumption rules out discontinuous changes in the mortality rate or prevalence of severe diagnoses in the Charlson index after any specific birthday. For healthcare utilization, one might think of institutional reasons for sudden jumps in the number of GP visits or hospitalizations. However, In Denmark there are no institutional differences in terms of healthcare costs or availability between an individual at age 59 and 60. Due to the universal hospital and GP coverage, there is no out-of-pocket expenses for neither a 59 nor a 60-year old. Also, there are no compulsory screenings or similar institutional health procedures which only apply for 60-year olds. The continuity assumption can be partly validated by examining jumps in the dependent variable at other ages than 60. If there are no such jumps, it is suggestive of the validity of the continuity assumption. Second, I assume that individuals cannot manipulate the assignment variable (age). Because age is predetermined, this assumption seems natural. Third, there must be no other confounding discontinuities in the explanatory variables at age 60. In Denmark, the only difference between a 59-year-old and a 60-year-old stems from the retirement legislation. Similar to the assumption of continuity in the dependent variable (apart from at age 60), the lack

[13]This assumption is in line with biological theories of aging as presented in e.g. Mitnitski et al. (2002)

of confounding discontinuities can be partly validated by performing tests on the observed variables. If there are any discontinuities in the explanatory variables (or other relevant observed variables) around the threshold, it points towards violations of this assumption. In the results section, I provide graphical and formal evidence showing that this is not a concern.

Given the assumptions, I can estimate the causal effect of retirement on health by comparing the (potential) jump in health or healthcare utilization at age 60, with the discontinuity in retirement takeup at age 60. Formally I do this by estimating a 2SLS model of the form:

$$\mathrm{H}_{it} = \beta_0 + \beta_1 \mathrm{R}_{it} + \mathrm{X}_{it}\gamma + \mathrm{F}(\mathrm{age}_{it}, \mathrm{Z}_{it}) + u_{it} \tag{5}$$

$$\mathrm{R}_{it} = \delta_0 + \delta_1 \mathrm{Z}_{it} + \mathrm{X}_{it}\omega + \mathrm{G}(\mathrm{age}_{it}, \mathrm{Z}_{it}) + \nu_{it} \tag{6}$$

where H_{it} is a measure of health, R_{it} a retirement dummy and X_{it} a vector of controls. I instrument retirement with $\mathrm{Z}_{it} = \mathbf{1}(\mathrm{age}_{it} \geq 60)$, a dummy for being above the age threshold, and let $\mathrm{F}(\cdot)$ and $\mathrm{G}(\cdot)$ be smooth functions of age. The vector of controls, X_{it}, is not required for identification but including controls may decrease the variance of the estimates. Consequently, I run regressions without controls in the main specification and regressions with controls in the subsequent robustness section. I focus attention on individuals close to the cutoff by estimating local regressions using the triangular kernel[14] and a bandwidth of 5. Therefore, the model formed by equation (5) and (6) is a local IV model (see e.g. Imbens and Lemieux

[14]As referenced in Fan and Gijbels (1996), the triangular kernel is optimal in a MSE sense for local polynomial regressions at boundary points. Boundary point properties are important because estimates at the boundary point (age 60) form the basis for this RD design. In the robustness section of this paper, I perform regression with the rectangular kernel as well.

(2008) or Hahn, Todd and Van der Klaauw (2001)) putting most weight on observations close to the cutoff (age 60). I model the smooth relationship between H_{it} and age as a second order polynomial in age, allowing for different coefficients on either side of the discontinuity as suggested in Lee and Lemieux (2010). I use age polynomials of the same degree in the first and the second stage ($\text{deg}(\text{F}(\cdot)) = \text{deg}(\text{G}(\cdot))$) but let the coefficients vary freely. For the second stage, the function of age is

$$\begin{aligned} \text{F}(\text{age}_{it}, \text{Z}_{it}) =& \eta_1(\text{age}_{it} - c) + \eta_2(\text{age}_{it} - c)^2 \\ &+ (\eta_3(\text{age}_{it} - c) + \eta_4(\text{age}_{it} - c)^2)\text{Z}_{it} \end{aligned} \tag{7}$$

where c is the age threshold (60). Z_{it} enters because I let the polynomial vary on either side of the threshold. Dropping the term involving Z_{it} in equation (7), corresponds to restricting the coefficients of the age polynomial to be identical across the threshold. The choice of $\text{F}(\cdot)$ is an important part of the specification, so I also estimate locally linear models (as opposed to the local quadratic models) in the robustness section. In addition to this specification check, I evaluate the impact of bandwidth and kernel choice by estimating model (5) with a bandwidth of 10 years and with a rectangular kernel. Further, I estimate the locally linear model with a rectangular kernel and a narrow 2 year bandwidth. Finally, I estimate model (5) in a specification with added covariates and in a FE-2SLS specification identifying the effect off the within-individual variation only. The results are very similar across all of these additional specifications.

In model (5), $\hat{\beta}_1$ is the estimate of the average treatment effect for compliers. Compliers are the individuals who retire at age 60 because the retirement legislation is much more favorable to a 60-year old compared to a 59-year old, but who would have continued to work had this not been the case. Knowing the average causal effect for these individuals is very important for policy recommendation purposes, because they are exactly the kind of individuals who will respond to changes in the statutory retirement age.

5.2 RD results

Figure 3 shows the retirement ratio by age with the superimposed local quadratic fit (on either side of the threshold) using a bandwidth of 5 years and a triangular kernel. The figure can be thought of as the graphical first-stage and shows the large discontinuous jump in the retirement ratio at age 60 - the earliest age of retirement for individuals without disabilities or other special circumstances. In addition to the large discontinuous jump at age 60, the age-trajectory in the retirement rate differs on either side of the age threshold. This difference is an artifact of the Danish retirement system, with only one substantial retirement option before age 60 (disability pension) and a handful of retirement options and incentives after age 60. The differential age trajectories highlights the importance of allowing the age-functions, $F(\cdot)$ and $G(\cdot)$, to vary on either side of the threshold.

Figure 4 shows the graphical reduced form evidence of the effect of retirement on health and healthcare utilzation. The figure displays no effects on the direct measures of health but small negative effects on healthcare usage. Specifically, the Charlson Index and the 1-year mortality rate evolve smoothly around the age 60 cutoff, whereas GP visits and hospitalizations drop after retirement. Because turning 60 does not cause everyone to retire (retirement takeup increases around

17 percentage points between age 59 and 60), the visible effect in figure 4 will be about 5 times smaller than the causal effect for compliers.

Table 3 shows estimates from local regressions (bandwidth of 5 years and triangular kernel) of health or healthcare utilzation onto retirement. Each column contains estimates for a specific variable and each panel represents a specific regression model. For all regressions in table 3, age is modeled with a quadratic polynomial on both sides of the age threshold (as presented in equation (7)). The OLS panel contains regressions of model (5) without use of instruments. Compared to the reform results, these estimates of the association between retirement and health show retirement to be associated with sharper declines in health and larger increases in healthcare utilzation. This is a consequence of the younger RD sample (aged around 60 compared to 65+ for the reform sample) and the fact that retirement before age 60 is driven by disability pension recipients[15]. The pure associations are presented as a benchmark for the 2SLS estimates.

The reduced form panel presents estimates from model (5) with an indicator for (age ≥ 60) instead of the retirement dummy. The reduced form panel of table 3 confirms the initial graphical evidence from figure 4 of no effects on the Charlson index or mortality, but negative effects on healthcare utilization measured by the number of GP visits and the hospitalization rate. The reduced form effects are small but statistically significant.

The 2SLS panel contains the estimates from model (5) instrumenting retirement with the age threshold (equation (6) as the first stage), i.e. the causal effect of retirement on health and healthcare utilization. The estimate from the Charlson

[15] Disability pension is more negatively associated with health than ordinary retirement because disability pension eligibility requires ill health or specific social conditions

index measure shows no significant effects of retirement. The point estimate and standard errors are small, ruling out both large negative and large positive effects.

Contrary to the zero effects on the Charlson index, retirement leads to a drop in the risk of being hospitalized of 1.1 percentage points out of a mean hospitalization rate of 15.4% - a drop of roughly 7%. Similarly, retirement leads to 0.35 fewer yearly GP visits, which corresponds to around 8% of the mean number of GP visits. The negative effects for healthcare utilization are very similar to the effects on healthcare utilization found in (Hallberg, Johansson and Josephson, 2015) (using a reform strategy) but smaller in magnitude than the effects in Eibich (2015) (using an RD strategy). Hallberg, Johansson and Josephson (2015) find the introduction of an early retirement offer to reduce the number of days in inpatient care with 35% over 5 years and that an additional year of retirement leads to an 8% reduction in the days in inpatient care. The Hallberg, Johansson and Josephson (2015) per-year effect sizes are comparable to the 7% reduction in the risk of being hospitalized found in the RD design of this study. Eibich (2015) examines both an indicator for hospital stay within a year and the number of GP visits within 3 months and finds that retirement leads to reductions of 31% and 26% of the average for hospital stays and GP visits, respectively. These effect sizes are large compared to those in this study even though measures should be quite comparable.

As the case with the Charlson index, there is no signifcant effect on mortality.

Comparing the estimates in the 2SLS panel with the associations in the OLS panel, we see that the strong associations between retirement and poor health/more healthcare utilization shown in the OLS panel, is driven purely by selection into retirement. The estimates in the 2SLS panel are causal, showing zero effects of retirement on the Charlson index and mortality and negative effects on healthcare utilization - meaning that retirement leads to less usage of the healthcare system. This highlights that a simple comparison of the health of retirees and non-retirees is not appropriate for empirical investigations of the causal effects of retirement on health.

5.3 RD heterogeneity

Due to the number of available observations, the RD design allows for analyses of effect heteregeneity. First, I investigate the effect of retirement on health and healthcare utilization for men and women. I do this by estimating models with the same specification as the main 2SLS model in table 3 separately for men and women. I estimate the models seperately for men and women to allow for the maximal amount of flexibility according to gender[16]. Table 4 presents the results. Comparing means of the dependent variables, we see that despite being in worse health than women, men are less frequent users of primary healthcare but are admitted to hospital to the same extent as women. A reason behind differences in GP visits but not in hospitalizations, could be that individuals have more control over the decision to visit the GP than over being admitted to the hospital. In such a case, differences in healthcare utilization behavior across gender will be most apparent from looking at the number of GP visits. Moving onto the effects of retirement, we see no significant effects of retirement on the Charlson index or mortality for neither men nor women. This is analogous to the main results. For the healthcare utilization variables, there

[16] Another option would be to estimate models similar to the main RD model, but including terms interacted with dummies for male. Such procedures are less flexible than separate estimation, unless all terms are interacted with the dummy (in which case the procedures coincide).

is a rather surprising pattern. For women, there is no significant effect on hospital admittance but rather large negative effects on the number of GP visits, with effects corresponding to about 13% of the mean number of visits for women. For men, the effect on hospital admittance is negative but there is no significant effect on GP visits. This decline in the risk of being hospitalized, corresponds to just above 11% of the mean hospitalization rate.

Table 5 provides additional evidence on effect heterogeneity, by showing the estimates split by income rank at age 55. The income rank is cohort specific, measuring relative income at age 55 within the respective cohort. This cohort specificity handles potential issues with inflation and cross-cohort differences. From the first rows in each panel, it is clear that individuals with higher incomes tend to be healthier (lower values of the Charlson index and lower 1-year mortality rates) and utilize the healthcare system to a smaller degree. From the coefficient on retirement, we see that retirement leads to a decrease in GP visits for individuals in the bottom of the distribution, but has no significant effect on GP visits for individuals with high incomes. The pattern is reversed for the risk of being hospitalized, with negative effects for high income individuals but no significant effects for low income individuals. This is similar to the results split by gender, with similarity between low-income individuals and women and high-income individuals and men. Because women tend to have lower labor market earnings than men, the similarity between estimates in table 4 and 5 could be driven by the same variable (e.g. income). As the case in table 4, there are no significant effects on the Charlson index or mortality, regardless of relative income rank at age 55

To investigate the income/gender differences further, table 6 shows the estimates split by gender for individuals in the bottom half of the income distribution while table 7 shows the corresponding gender splits for individuals in the top of the income distribution. Table 6 shows that for women in the bottom of the income distribution, retirement leads to fewer GP visits. There are no significant effects for the other measures of health or healthcare utilization for low income women and no effects for any of the measures for low income men. Table 7 shows the corresponding estimates for high income individuals. For these individuals, the gender differences are qualitatively similar to the gender differences in table 4 (which disregards the income dimension). Together, the estimates in table 6 and 7 suggest that income is not the sole driver of the differences between men and women because the gender differences are qualitatively similar when income is taken into account. Consequently, women visit the GP to a lesser extent after retirement while men go to the hospital to a lesser extent after retirement. For both genders, retirement has no effect on the Charlson index or motality.

5.4 RD robustness

This subsection performs a battery of checks to evaluate the robustness of the main results, starting from traditional checks of RD designs and moving towards more elaborate alternative specification.

Other discontinuities Because individuals cannot manipulate their age in the Danish registers, we would not expect any discontinuities in the number of observation close to the cutoff due to assignment variable manipulation. However, migration, cohort size differences or mortality could cause discontinuities in the density of observations, comprimising clean identification. To investigate whether differences in the number of observation could bias the results, figure C.1 depicts the number of non-missing retirement observations by age. The lack of discontinuities around age 60 in figure C.1, suggests that the number of observations should not be a cause of concern.[17]

Turning to discontinuities in other covariates, figure C.2 plots the average birth year, share with a college degree, share of married individuals and share of men by age. The top-left subfigure shows the average year of birth declining with age - a consequence of the age and year restrictions. There is no sudden jumps in average cohort - which suggests that cohort differences are not likely to affect the main estimates. The share with a college degree, share married and share male shows no discontinuities around age 60 but generally decline with age due to a mix of mortality and cohort effects. As long as there are no discontinuities around age 60, these declines do not invalidate the RD design. Figure C.3 zooms in on income, showing a discontinuity around age 60. This discontinuity does not violate the necessary assumptions because retirement is expected to be followed by decreases in income. The magnitude of the jump in income is relatively modest, which is probably due to the end-of-year measures of age and retirement and the annual measure of income. If an individual turns 60 in November and retires immediately, I will observe her as retired at age 60 at the end of the year. However, because she has been working and receiving wages for the majority of the year. I will not see huge declines in her

[17] There is an increase in the number of observations going from age 50 to 51, which is a mechanical effect arising due to the sample and year restrictions. Because only individuals born before 1960 are observed, we don't observe any 50-year-olds in 2010 resulting in one year less with observations at age 50 (compared to all older ages).

yearly earnings [18]. Table C.1 contains the regression equivalent of the the graphical evidence. Specifically, table C.1 contains 2SLS estimates with the same modelling choices as in the main table 3 with income, college, married, male and cohort as the dependent variables. Table C.1 shows the expected significant decline in income post retirement as well as no effects on marriage, share male or cohort. Moreover, the table shows that retirement leads to a very small increase of 0.3 percentage points in the probability of having a college degree. This was not visible in figure C.2. To investigate this further, table C.2 shows results for similar regressions, but on a balanced sample of individuals who are observed at all ages from 55-65. By balancing the sample, I eliminate cohort selection effects. The regressions on the balanced sample shows no effects of retirement on college or any of the other covariates[19], except for income which declines as a direct consequence of retirement. To sum up, the graphical investigations into the RD design robustness does not reveal any unexpected effects of retirement on covariates. This conclusion extends to the evidence from regressions on both the unbalanced and balanced sample, except for the tiny effects on college education in the unbalanced sample. The effects on college vanish once I restrict estimation to the balanced sample, pointing towards cohort selection as the main cause behind the effects in the unbalanced sample. In the subsequent robustness analyses, I show that including cohort or year FE's and estimating the effects off the within individual variation in a FE-2SLS specification does not change the main results - an important indication that selection does not

[18]There is a second effect as a result of partial retirement in which individuals continue to work part-time after their official retirement. Unfortunately, there is no distinction between full and partial retirement in the Danish registers so this aspect cannot be explored further.

[19]There are no standard errors on the estimated coefficients in the regressions for male or cohort, because these dependent variables are time-invariant, resulting in rank deficient covariance matrices.

invalidate the main results.

Alternative specifications, kernels and bandwidths I continue the robustness evaluation of the main results, by performing regressions of health and retirement varying the bandwidth, kernel choice and model specification. Results for each dependent variable are presented in table C.3 to C.6. In each table, the first column contains OLS estimates from model (5) with age specification (7), identical to the estimates in the OLS panel of table 3. The second column contains estimates from model (5) with age specification (7) instrumenting retirement with the threshold dummy, equivalent to the estimates in the 2SLS panel of table 3. The first two columns are included for ease of comparability. The third column of each table replicates the main 2SLS results but increases the bandwidth from 5 to to 10 years. The fourth column uses a rectangular kernel instead of the triangular kernel. The fifth column models age as locally linear instead of locally quadratic (setting $\eta_2 = \eta_4 = 0$ in equation (7) as well as the corresponding coefficients in $G(\cdot)$). The sixth column adds year dummies, dummies for income decile at age 55, dummies for male, college education and married. Adding year dummies is equivalent to adding cohort dummies (because age is also included) and ensures that cohort or year specific effects do not bias the estimates. These estimates relate to the earlier discussion of cohort selection effects. The seventh column of table C.3, C.4 and C.5 adds individual specific FE's to the specification in column (2) and thus handles all time invariant covariates (observed and unobserved). Table C.6 does not contain the FE estimates because of the cross-sectional nature of mortality with no useful within-individual variation.

Table C.3 contains the estimates for the Charlson index. Looking across the model specifications in columns (2) to (6) we see that the coefficient on retirement remains small in magnitude and insignificant. This shows that neither the choice of kernel, bandwidth or inclusion of controls changes the results. This is well in line with the previously presented evidence on health outcomes and distribution of covariates around age 60. Column (7) contains the FE specification, showing significant negative effects of retirement on the Charlson index (positive effects on health). The significant coefficient in the FE-2SLS specification hinges on the exact FE-2SLS specification and is not robust with respect to modelling choices. Because all remaining estimates show no significant effect of retirement on the Charlson index, I will not base any conclusions on the coefficient from the FE-2SLS Charlson specification.

Table C.4 shows the specification checks for the hospitalization dummy. The results are very similar across choice of bandwidth, controls, FE's and kernel. The effects of retirement on the risk of being hospitalized vary from -1.0 percentage points in the FE specification to -1.3 percentage points in the specification with the rectangular kernel. The remaining results are all within this band. The high levels of comparability across models are signs of the robustness of the main results.

Results from the robustness check of the results for GP visits are depicted in table C.5. The results are similar across specifications, with the effects of retirement varying from -0.47 GP visits per year in the FE specification to -0.26 GP visits per year in the case with the rectangular kernel. The main estimate presented in column (2) is placed between these two estimates. Comparing these results with those for hospitalzations in table C.4, we see that there is no systematic tendency for the FE coefficients to be largest in magnitude or for the rectangular kernel coefficients to be smallest in magnitude. Furthermore, point estimates in column (3)-(7) in both table C.4 and C.5 are included in the 95% confidence interval around the main results coefficients in the second column.

Table C.6 presents the robustness results for 1-year mortality. Looking across columns (2) to (6), we see that the point estimates (along with the standard errors) do not change with inclusion of controls, increases in bandwidth or changed kernel. As in the main results table, the point estimates are small and slightly positive in all specifications, but not significant. Again, this points toward robustness of the main results.

Placebo threshold As a final test of the RD design, I implement a placebo procedure testing whether I can find any significant effects on health using thresholds other than age 60. If I can find effects on health for age thresholds without any retirement discontinuities, it suggests that something other than retirement differences may be driving the results. It seems natural to perform the placebo tests using age 59 as the placebo threshold, using the same bandwidth as in the main specification. However, in that case we are in danger of misspecifying the smooth age function and thus (falsely) rejecting effects in the placebo case. To see why, return to figure 3 showing the retirement takeup by age with the clear discontinuity in retirement at age 60. By estimating the model using a placebo threshold at age 59, we would have to include observations at age 60. This is problematic, because these observations are affected by the large discontinuity at age 60. To avoid using observations which are affected by the actual discontinuity, I choose a placebo threshold age of 57 and limit the bandwidth to 2 years. Further, I use the simple rectangular kernel and model age as locally linear (as opposed to quadratic) on either side of the threshold. For completeness, I also run the regressions with the true age 60 threshold which will serve as the baseline for the placebo regressions and a further robustness check of the main results.

Table C.7 shows the regression estimates for the baseline and placebo threshold for each of the dependent variables. The estimates in the odd columns play the dual role of baselines for the placebo test as well as robustness checks of the main results (narrow 2-year bandwidth, rectangular kernel and linear specification). The estimates in the odd columns should therefore be similar to the estimates in the bottom panel of main table 3. The estimates in the odd columns are very similar to the estimates in the main specification, with no significant effects of retirement on the Charlson comorbidity index or mortality but significant negative effects on hospitalizations and the number of GP visits. This again confirms the robustness of the main results. The estimates in the even columns are the placebo estimates, of which we expect weak instruments and insignificant estimates on retirement. The even columns in table C.7 shows that none of the placebo estimates are statistically significant or have the necessary first-stage significance to rule out weak instrument issues. This is not only visible in the first stage F-statistic reported in the bottom of the table, but also from looking at the estimated coefficients which are large and unprecisely estimated. In conclusion, all estimates in table C.7 suggest validity of the RD design.

6 Discussion

This paper estimates the causal effect of retirement using two different empirical strategies: a reform induced change in the old age pension age and a large discontinuity in retirement takeup at age 60. In the reform strategy, the effect of retirement on health is not estimated with enough precision to rule out effects being either positive or negative. However, the graphical evidence suggests that if there are any effects, they are small. In contrast to the reform strategy, the RD design has the statistical power to find, and rule out, small effects of retirement on health. Indeed, the RD design shows negative effects on healthcare utilization with effect sizes corresponding to a drop in healthcare utilization of 7-8% of average utilization. Further, the RD results rule out important effects on the Charlson index and mortality. Since retirement leads to a drop in healthcare utilization, retirement generates small positive effects on governmental healthcare spending. However, these effects are negligible in comparison with the negative effects on government budgets stemming from the retirement-induced decreases in income taxes. A rough calculation for GP visits, disregarding inflation, taxes on retirement savings and other complications, highlights the relative magnitudes of public healthcare savings and income taxation losses from retirement. With retirement leading to 0.35 fewer GP visits per year (table 3) and an average expenditure per consultation of 273 DKK[20], retirement leads to savings in public GP expenditure of around 95 DKK per person in the year of retirement. In terms of losses from declines in income tax revenue, retirement leads

[20]Data on public healthcare spending and utilization from Statistics Denmark (Statistics Denmark, 2017) for 2006. The total expenditure on GP's for 60-year-olds in 2006 was 93.240 million DKK and the total number of GP consultations for 60-year olds were 342,046, yielding average cost per consultation of $\simeq$ 273 DKK.

to a drop in income of about 54,000 DKK (table C.1) corresponding to losses in tax revenue of roughly 23,000 DKK with a 42% tax rate[21] in the year of retirement. Even though the primary sector only accounts for about 10% of the total Danish healthcare expenditures (see Christensen, Gørtz and Kallestrup-Lamb (2016)) this rough sketch shows that the positive effects on public healthcare spending are small compared to the losses in revenue from income taxes.

A main limitation of the RD design is that it only allows for estimation of short-term effects. Because identification stems from the discontinous jump in retirement propensity at age 60, the RD design can only be used to identify differences in health just around the cutoff. This restriction is especially relevant for measures of mortality and severe diagnoses and less relevant for measures of healthcare utilization. For example, we might not expect to see individuals developing diabetes after a year of retirement, whereas visits to the GP could be more responsive. This could explain why there is no statistically significant effects on the Charlson index or mortality, but significant effects on the number of GP visits and the risk of hospitalization.

Despite of the supposedly sluggish nature of the evolution of mortality and severe diagnoses, there is a rationale for looking at the Charlson index and mortality. Even though diabetes (presumably) does not develop fully over the short course of a year,[22] an overall change in health, health investments or well-being may lead to differences in the diabetes propensity in the affected population. After retire-

[21]The Danish Ministry of Taxation (Ministry of Taxation, Denmark, 2017) lists the average marginal tax rate for individuals facing the lowest marginal tax rates to 42.9% in 2004 and 42.6% in 2008. Individuals also pay taxes upon payout of retirement savings, but these will be taxed upon payout regardless on payout timing and can thus be disregarded. I do not consider differential taxation of pensions, retirement means-testing or other complexities in the Danish tax system.

[22]Unless retirement has truly drastic impacts on behavior

ment, the marginal diabetes patient may experience a delayed or accelerated need for medical diabetes treatment through changes in health behavior or other aspects of health. The empirical strategies should be able to find such effects if they are present in the data.

With the data at hand, it is a difficult task to disentangle the various mechanisms through which retirement leads to decreases in healthcare utilization. The decline in healthcare utilization may be caused by decreases in individual health investments (visiting the doctor less frequently for preventive health examinations or advice), improvements in health (and therefore decreases in the need to see a doctor for treatment of specific conditions) or a mix of these. Furthermore, even *increases* in health investments along an unobserved dimension such as exercise (as found in e.g. Kämpfen and Maurer (2016)) affecting overall health, could reduce the need for medical treatment and thus lead to less usage of public healthcare. The results are therefore inconclusive with respect to the mechanisms driving the decrease in healthcare utilization. However, in terms of the effects on public health spending, the effects are small but positive (i.e. retirement decreases public healthcare spending).

Given the small magnitude of the effects found in this study, it is instructive to consider under which (if any) circumstances one would expect *large* effects of retirement on health or healthcare utilization. One example is income. If income has a strong influence on the availability of healthcare, one might expect large negative effects of retirement on healthcare utilization due to the drop in income associated with retirement. In Denmark, the universal healthcare coverage means that income should not have any direct effect on health or healthcare utilization. Therefore, the effects from income are probably quite small for Danes compared to individuals from countries with dissimilar healthcare systems but comparable to individuals from countries with simlar healthcare systems. One might also expect large effects for individuals with working conditions with strongly adverse health effects (such as arduous manual labor or large amounts of stress). In this dimension, Denmark is probably comparable to most developed countries. To conclude, there is no a priori reason to believe that the causal effect of retirement on health should be markedly different when it is estimated on Danish data compared to estimates from countries with similar healthcare systems. Perhaps, individual health behavior and health status simply does not change too drastically after retirement.

References

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

Behncke, Stefanie. 2012. "Does retirement trigger ill health?" *Health economics*, 21(3): 282–300.

Bingley, Paul, and Peder J Pedersen. 2011. "Does early retirement keep you healthu and extend your life?" *Preliminary version) The Danish National Center for Social Research, Copenhagen School of Economics and Management, Aarhus University*.

Bloemen, Hans, Stefan Hochguertel, and Jochem Zweerink. 2017. "The causal effect of retirement on mortality: Evidence from targeted incentives to retire early." *Health Economics*, , (January): 1–15.

Bloom, David E, Somnath Chatterji, Paul Kowal, Peter Lloyd-Sherlock, Martin McKee, Bernd Rechel, Larry Rosenberg, and James P Smith. 2015. "Macroeconomic implications of population ageing and selected policy responses." *The Lancet*, 385(9968): 649–657.

Buckles, Kasey S, and Daniel M Hungerman. 2013. "Season of birth and later outcomes: Old questions, new answers." *Review of Economics and Statistics*, 95(3): 711–724.

Christensen, Bent Jesper, Mette Gørtz, and Malene Kallestrup-Lamb. 2016. "Medical spending in Denmark." *Fiscal Studies*, 37(3-4): 461–497.

Coe, Norma B, and Gema Zamarro. 2011. "Retirement effects on health in Europe." *Journal of health economics*, 30(1): 77–86.

Contoyannis, Paul, Jeremiah Hurley, Paul Grootendorst, Sung-Hee Jeon, and Robyn Tamblyn. 2005. "Estimating the price elasticity of expenditure for prescription drugs in the presence of non-linear price schedules: an illustration from Quebec, Canada." *Health Economics*, 14(9): 909–923.

Dave, Dhaval, Inas Rashad, and Jasmina Spasojevic. 2008. "The Effects of Retirement on Physical and Mental Health Outcomes." *Southern Economic Journal*, 497–523.

Dee, Thomas S, and Hans Henrik Sievertsen. 2015. "The Gift of Time? School Starting Age and Mental Health." National Bureau of Economic Research.

Eibich, Peter. 2015. "Understanding the effect of retirement on health: mechanisms and heterogeneity." *Journal of health economics*, 43: 1–12.

Fan, Jianqing, and Irene Gijbels. 1996. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66.* Vol. 66, CRC Press.

Godard, Mathilde. 2016. "Gaining weight through retirement? Results from the SHARE survey." *Journal of health economics*, 45: 27–46.

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

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica*, 69(1): 201–209.

Hallberg, Daniel, Per Johansson, and Malin Josephson. 2015. "Is an early retirement offer good for your health? Quasi-experimental evidence from the army." *Journal of health economics*, 44: 274–285.

Heller-Sahlgren, Gabriel. 2017. "Retirement Blues." *Journal of Health Economics*.

Imbens, Guido W, and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of econometrics*, 142(2): 615–635.

Johansen, Martin Berg, and Christian Fynbo. 2011. "Comparison of Charlson comorbidity index with SAPS and APACHE scores for prediction of mortality following intensive care." *Clinical epidemiology*, 3: 203–211.

Johnston, David W, and Wang-Sheng Lee. 2009. "Retiring to the good life? The short-term effects of retirement on health." *Economics Letters*, 103(1): 8–11.

Kämpfen, Fabrice, and Jürgen Maurer. 2016. "Time to burn (calories)? The impact of retirement on physical activity among mature Americans." *Journal of health economics*, 45: 91–102.

Lee, David S, and Thomas Lemieux. 2010. "Regression discontinuity designs in economics." *Journal of economic literature*, 48(2): 281–355.

Ministry of Taxation, Denmark. 2017. Marginal tax rates. http://www.skm.dk/skattetal/statistik/tidsserieoversigter/marginalskatteprocenter-1993-og-1998-2017 [21st of March 2017].

Mitnitski, Arnold B, Alexander J Mogilner, Chris MacKnight, and Kenneth Rockwood. 2002. "The accumulation of deficits with age and possible invariants of aging." *The Scientific World Journal*, 2: 1816–1822.

Moore, Sophie E, Timothy J Cole, Elizabeth ME Poskitt, Bakary J Sonko, Roger G Whitehead, Ian A McGregor, and Andrew M Prentice. 1997. "Season of birth predicts mortality in rural Gambia." *Nature*, 388(6641): 434–434.

Motegi, Hiroyuki, Yoshinori Nishimura, and Masato Oikawa. 2016. "What Explains the Difference in the Effect of Retirement on Health?: Evidence from Global Aging Data."

Nielsen, Torben Heien. 2016. "The Relationship Between Self-Rated Health and Hospital Records." *Health economics*, 25(4): 497–512.

Shaffer, Juliet Popper. 1995. "Multiple hypothesis testing." *Annual review of psychology*, 46(1): 561–584.

Statistics Denmark. 2016. StatBank Denmark. http://www.statistikbanken.dk [02 January 2017].

Statistics Denmark. 2017. StatBank Denmark. http://www.statistikbanken.dk [21st of March 2017].

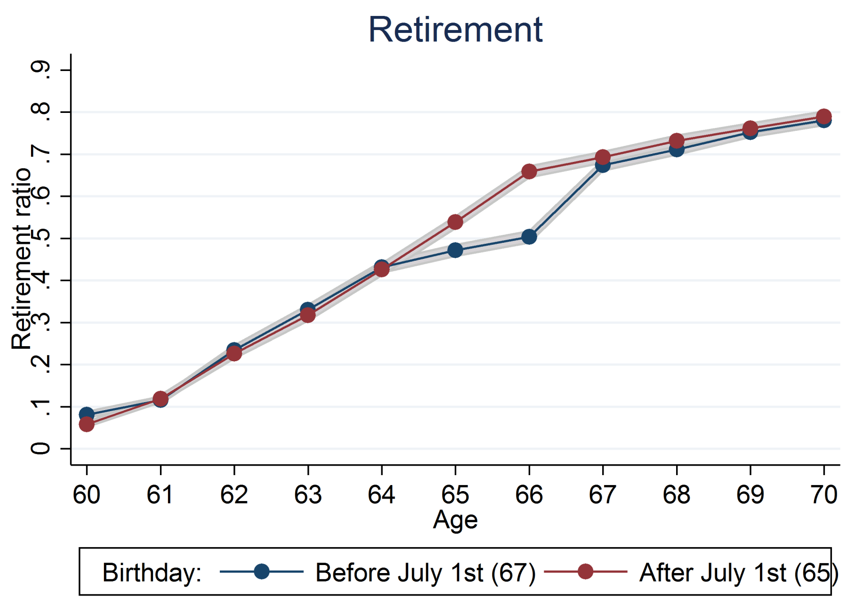
Whiteford, Peter, and Edward Whitehouse. 2006. "Pension challenges and pension reforms in OECD countries." *Oxford review of economic policy*, 22(1): 78–94.

Table 1: Summary Statistics

	Full/RD sample		Reform sample	
	Mean	SD	Mean	SD
Age 55				
Year	1996.2	8.83	1994	0
Male	0.50		0.54	
Has college degree	0.19		0.32	
Business owner	0.15		0.26	
In labor force	0.85		0.98	
Receives DIP	0.11		0	
Total earnings	194,583	596,194	221,391	327,745
Married	0.73		0.82	
Charlson	0.074	0.45	0.039	0.33
Hospitalized	0.15		0.089	
GP visits	4.06	5.19		
1-year mortality rate	0.0075		0.0075	
Age 60				
Eligible for ERP	0.71		0	
Retired	0.35		0.070	
Charlson	0.10	0.53	0.081	0.48
Hospitalized	0.15		0.15	
GP visits	4.31	5.19		
1-year mortality rate	0.012		0.0081	
Age 65				
Retired	0.67		0.50	
Charlson	0.14	0.62	0.12	0.62
Hospitalized	0.17		0.16	
GP visits	5.06	5.66		
1-year mortality rate	0.021		0.011	
Age 70				
Retired	0.85		0.78	
Charlson	0.17	0.67	0.29	0.89
Hospitalized	0.19		0.20	
GP visits	6.25	5.99	5.79	5.69
1-year mortality rate	0.029		0.016	
Individuals aged 55	1,978,936		10,986	
Total observations	86,447,912		220,743	

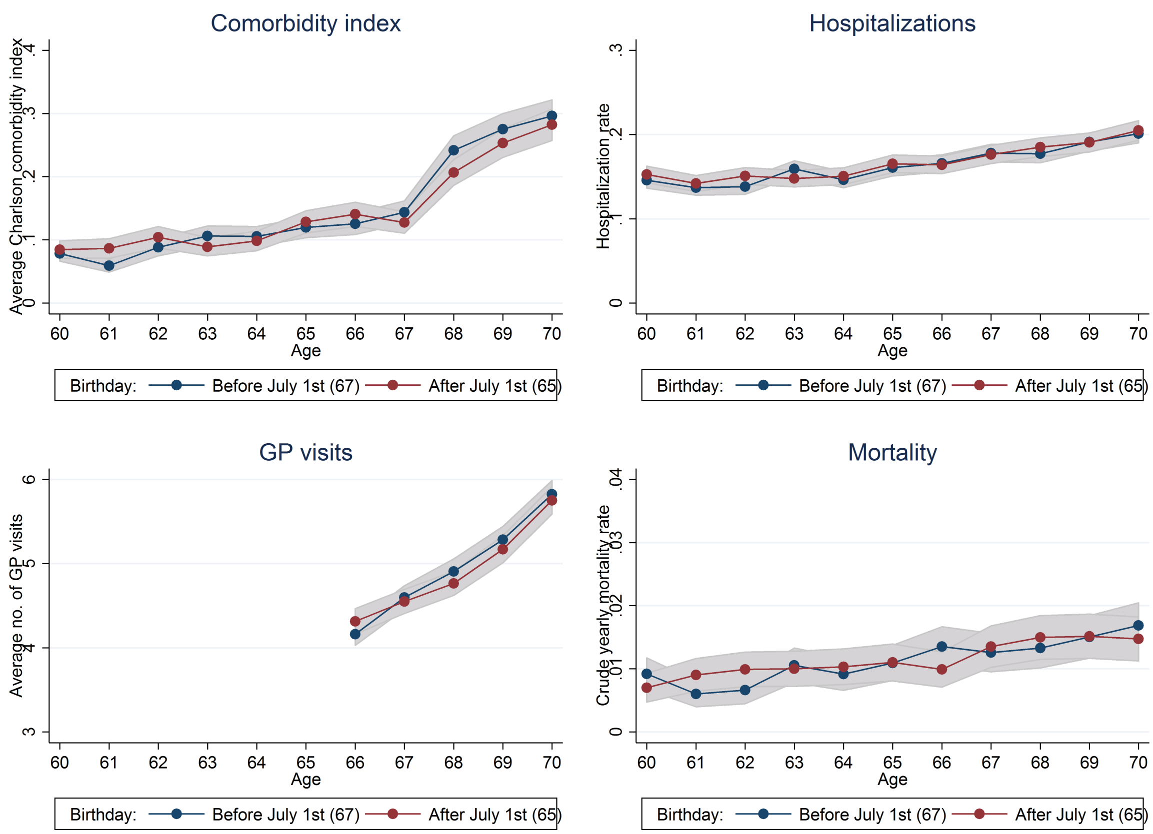
Summary statistics for the full sample (used in the RD strategy) and the sample used in the reform strategy. Earnings are nominal DKK (approximately 6.9 DKK/USD over the period). Years: 1980-2010 except for GP visits which are not available before 2006. Full sample: individuals born before 1960. Reform sample: 1939 cohort, individuals eligible for ERP excluded, Individuals receiving disability pension before age 60 excluded. Individuals excluded in the year of death.

Figure 1: Retirement takeup by birthday threshold



Retirement ratio conditional on age and birthday with 95% confidence intervals as shaded area. Sample: cohort 1939, individuals eligible for ERP excluded, individuals receiving disability pension before age 60 excluded, individuals excluded in the year of death.

Figure 2: Health and healthcare utilization by birthday threshold



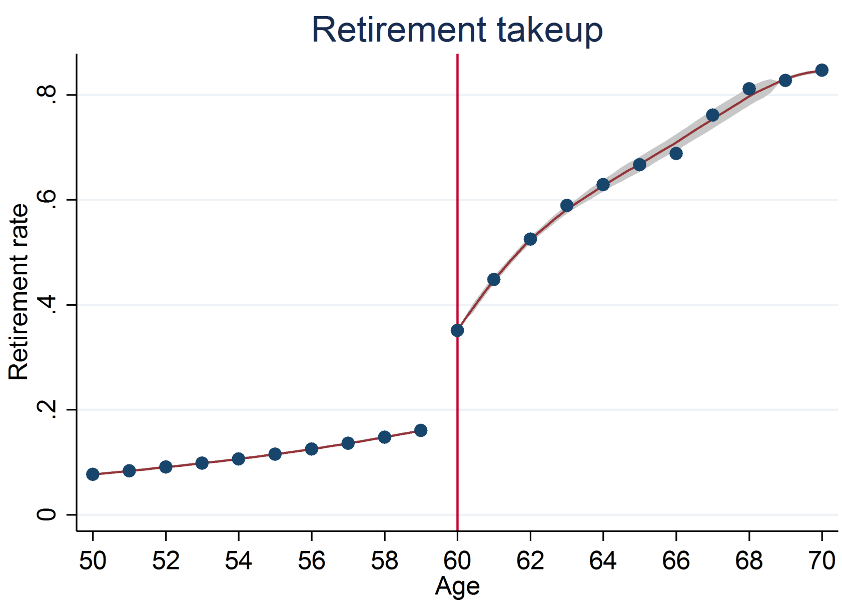
Mean values of the health or healthcare utilization variable conditional on age and birthday with 95% confidence intervals as shaded area. Sample: cohort 1939, individuals eligible for ERP excluded, individuals receiving disability pension before age 60 excluded, individuals excluded in the year of death.

Table 2: Reform strategy: main results

	(1) Charlson	(2) Hospitalized	(3) GP visits	(4) Mortality
Mean of dep. var	0.194	0.180	4.929	0.013
OLS				
Retired	0.080***	0.017***	0.700***	0.006***
	(0.009)	(0.004)	(0.087)	(0.001)
Reduced form				
$\mathbf{1}$(birthday $\geq$ July 1st	-0.010	0.003	-0.038	-0.000
	(0.010)	(0.004)	(0.086)	(0.001)
2SLS				
Retired	-0.216	0.056	-0.883	-0.008
	(0.205)	(0.086)	(2.000)	(0.020)
First stage F-stat	42.6	42.6	33.3	42.4
Individuals	9,924	9,924	9,709	9,908
Observations	57,418	57,418	45,865	57,334

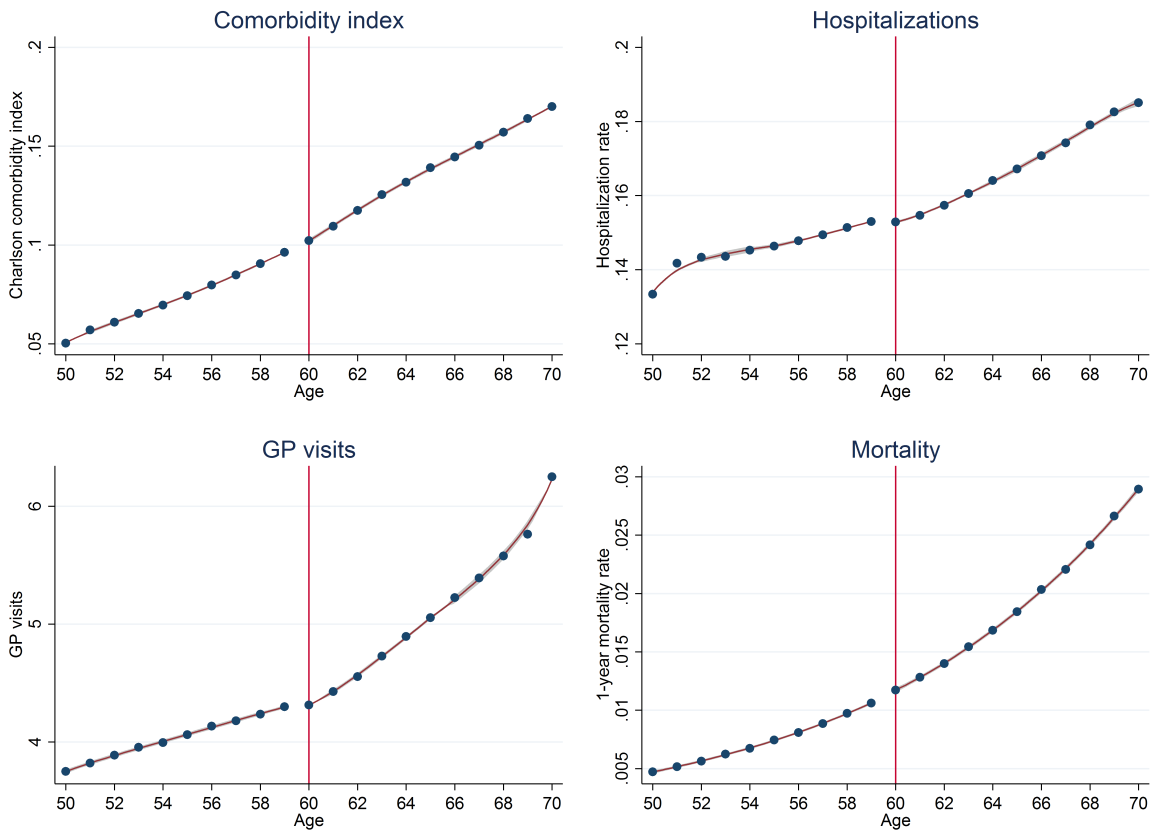
Each column contains estimates for a specific dependent variable and each panel represents a specific model. In all cases, controls are full sets of age and income decile dummies and dummies for male, college and married.
The OLS panel contains estimated coefficients on the retirement variable from regressions of the dependent variable onto a dummy for retirement and controls.
The reduced form panel contains estimated coefficients on the treatment indicator from regressions of the dependent variable onto a dummy for (birthday $\geq$ July 1st 1939) and controls.
The 2SLS panel contains estimated coefficients on retirement from 2SLS regressions of the dependent variable onto a dummy for retirement and controls, instrumenting retirement with the birth-date threshold.
Estimated for the 1939 cohort, age 65 to 70. Controls: full sets of age and income decile dummies, dummy for male, college and married. Standard errors clustered on the individual level in parentheses.
$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Figure 3: Retirement takeup by age



Retirement ratio by age. Lines show the local quadratic fit using a triangular kernel and a bandwidth of 5 years with 95% confidence intervals as shaded area. Sample: full/RD sample.

Figure 4: Health and healthcare utilization by age



Mean values of health or healthcare utilization variables by age. Lines show the local quadratic fit using a triangular kernel and a bandwidth of 5 years with 95% confidence intervals as shaded area. Sample: full/RD sample.

Table 3: RD strategy: main results

	(1) Charlson	(2) Hospitalized	(3) GP visits	(4) Mortality
Mean of dep. var	0.103	0.154	4.388	0.012
OLS				
Retired	0.089***	0.067***	1.678***	0.013***
	(0.001)	(0.000)	(0.013)	(0.000)
Reduced form				
$\mathbf{1}(\text{age} \geq \mathbf{60})$	-0.000	-0.002**	-0.060***	0.000
	(0.001)	(0.001)	(0.016)	(0.000)
2SLS				
Retired	-0.002	-0.011**	-0.348***	0.001
	(0.005)	(0.004)	(0.094)	(0.001)
First stage F-stat	215,928	215,928	33,478	223,537
Individuals	2,410,693	2,410,693	987,030	2,385,218
Observations	16,453,333	16,453,333	3,745,435	16,248,227

Each column contains estimates for a specific dependent variable and each panel represents a specific model. All models are local with triangular kernel and bandwidth of 5 years.
The OLS panel contains estimated coefficients on the retirement variable from local linear regressions of the dependent variable onto a dummy for retirement and a quadratic polynomial in age on either side of the threshold (age 60).
The reduced Form panel contains estimated coefficients on the treatment indicator from local linear regressions of the dependent variable onto a dummy for (age $\geq$ 60) and a quadratic polynomial in age on either side of the threshold (age 60).
The 2SLS panel contains estimated coefficients on retirement from local 2SLS regressions of the dependent variable onto a dummy for retirement and a quadratic polynomial in age on either side of the threshold (age 60), instrumenting retirement with the age threshold (age $\geq$ 60).
Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: RD strategy: heterogeneity by gender

	(1) Charlson	(2) Hospitalized	(3) GP visits	(4) Mortality
		Women		
Mean of dep. var	0.097	0.152	4.691	0.009
Retired	-0.008	-0.006	-0.610***	0.001
	(0.006)	(0.005)	(0.105)	(0.001)
First stage F-stat	125,259	125,259	24,211	129,708
Individuals	1,215,008	1,215,008	496,785	1,205,207
Observations	8,352,426	8,352,426	1,908,510	8,271,994
		Men		
Mean of dep. var	0.109	0.156	4.074	0.015
Retired	0.004	-0.017**	0.097	0.001
	(0.008)	(0.007)	(0.184)	(0.002)
First stage F-stat	91,195	91,195	10,173	94,262
Individuals	1,195,685	1,195,685	490,245	1,180,011
Observations	8,100,907	8,100,907	1,836,925	7,976,233

Heterogeneity analysis by gender.
Each column contains estimates for a specific dependent variable. Model specification is identical to the specification in the 2SLS panel of table 3.
Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: RD strategy: heterogeneity by income

	(1) Charlson	(2) Hospitalized	(3) GP visits	(4) Mortality
	Bottom half income			
Mean of dep. var	0.128	0.176	5.053	0.014
Retired	0.006	-0.006	-0.382***	0.001
	(0.007)	(0.005)	(0.114)	(0.002)
First stage F-stat	106,294	106,294	20,049	115,226
Individuals	945,512	945,512	482,067	934,361
Observations	6,880,827	6,880,827	1,820,290	6,781,277
	Top half income			
Mean of dep. var	0.090	0.149	3.746	0.008
Retired	-0.006	-0.014*	-0.261	0.002
	(0.007)	(0.006)	(0.165)	(0.002)
First stage F-stat	117,314	117,314	18,331	118,114
Individuals	944,508	944,508	496,680	939,670
Observations	7,022,007	7,022,007	1,901,667	6,961,291

Heterogeneity analysis by income.
Each column contains estimates for a specific dependent variable and with a panels representing each half of the income distribution at age 55. Model specification is identical to the specification in the 2SLS panel of table 3.
Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: RD strategy: heterogeneity by gender, low income

	(1) Charlson	(2) Hospitalized	(3) GP visits	(4) Mortality
	Women with low income			
Mean of dep. var	0.111	0.165	5.129	0.010
Retired	0.000	-0.006	-0.538***	0.001
	(0.008)	(0.006)	(0.121)	(0.002)
First stage F-stat	83,140	83,140	17,066	88,189
Individuals	624,111	624,111	308,928	619,019
Observations	4,675,168	4,675,168	1,187,179	4,626,179
	Men with low income			
Mean of dep. var	0.163	0.198	4.912	0.022
Retired	0.022	-0.005	0.018	-0.000
	(0.015)	(0.011)	(0.267)	(0.004)
First stage F-stat	24,908	24,908	3,955	28,253
Individuals	321,401	321,401	173,139	315,342
Observations	2,205,659	2,205,659	633,111	2,155,098

Heterogeneity analysis by gender for individuals in the bottom half of the income dsitribution at age 55.
Each column contains estimates for a specific dependent variable. Model specification is identical to the specification in the 2SLS panel of table 3.
Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: RD strategy: heterogeneity by gender, high income

	(1) Charlson	(2) Hospitalized	(3) GP visits	(4) Mortality
	Women with high income			
Mean of dep. var	0.087	0.155	3.955	0.006
Retired	-0.015	-0.005	-0.800***	-0.000
	(0.011)	(0.009)	(0.211)	(0.002)
First stage F-stat	50,866	50,866	8,964	51,192
Individuals	323,903	323,903	184,270	322,734
Observations	2,359,308	2,359,308	710,951	2,344,626
	Men with high income			
Mean of dep. var	0.091	0.146	3.622	0.010
Retired	-0.000	-0.020*	0.218	0.003
	(0.010)	(0.009)	(0.246)	(0.002)
First stage F-stat	67,111	67,111	9,480	67,577
Individuals	620,605	620,605	312,410	616,936
Observations	4,662,699	4,662,699	1,190,716	4,616,665

Heterogeneity analysis by gender for individuals in the top half of the income dsitribution at age 55. Each column contains estimates for a specific dependent variable. Model specification is identical to the specification in the 2SLS panel of table 3.
Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix A

Table A.1: ERP eligiblity and takeup

	60		62		64	
On ERP:	0	1	0	1	0	1
Not eligible	0.97	0.03	0.94	0.06	0.88	0.12
Eligible	0.63	0.37	0.43	0.57	0.29	0.71

Tabulation of ERP eligibility and ERP takeup at age 60-64 for the 1939 cohort. Each cell contains the ERP takeup rate as a percentage of all individuals in each row (split by age). 68.4% of the 1939 cohort were eligible for ERP and the cohort size was 55,648 at age 60.

Appendix B: reform robustness

Table B.1: Reform strategy: balance table

1939 birthmonth:	Jan.-Jun.		Jul.-Dec.		Test of equality
	Mean	SD	Mean	SD	p
Male	0.53		0.54		0.53
Has college degree	0.31		0.32		0.78
Business owner	0.26		0.25		0.54
In labor force	0.98		0.98		0.84
Total earnings	225,938	383,657	216,757	254,380	0.14
Married	0.82		0.81		0.16
Charlson index value	0.037	0.32	0.040	0.34	0.64
Hospitalized	0.088		0.089		0.76
1-year mortality rate	0.0064		0.0088		0.15
Observations	5,738		5,328		11,066

Summary statistics and balancing test at age 55 for the sample used in the reform strategy split by birthday. Test of equality presents the p-values for 2-sided tests of equal means or proportions. Reform sample: 1939 cohort, individuals eligible for ERP excluded, Individuals receiving disability pension before age 60 excluded. Individuals excluded in the year of death.

Table B.2: Reform strategy: Charlson index

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
Retired	0.079***	-0.215	-0.216	-0.216
	(0.008)	(0.205)	(0.205)	(0.205)
Age	-0.622***	-0.245	-0.231	
	(0.140)	(0.306)	(0.305)	
Age squared	0.005***	0.002	0.002	
	(0.001)	(0.002)	(0.002)	
Male	0.076***	0.017	0.026	0.026
	(0.010)	(0.042)	(0.034)	(0.034)
College	-0.030**	-0.031**	-0.022	-0.022
	(0.010)	(0.010)	(0.013)	(0.013)
Married	-0.014	-0.011	-0.004	-0.004
	(0.011)	(0.012)	(0.012)	(0.012)
Full age dummies	No	No	No	Yes
Income decile dummies	No	No	Yes	Yes
First stage F-stat		41.4	42.6	42.6
Individuals	10,337	10,337	9,924	9,924
Observations	59,566	59,566	57,418	57,418

Estimated coefficients from regressions of the Charlson index onto retirement. Column (1) contains estimates from model (3) without use of instrument (eq. (4) is not used) Column (2)-(4) contain estimates from model (3) instrumenting retirement with a dummy for (birthday $\geq$ July 1st 1939) (eq. (4) as the first stage). Estimated for the 1939 cohort, age 65 to 70. Standard errors clustered on the individual level in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.3: Reform strategy: hospitalization

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
Retired	0.017***	0.044	0.056	0.056
	(0.004)	(0.085)	(0.086)	(0.086)
Age	-0.113	-0.147	-0.139	
	(0.080)	(0.135)	(0.135)	
Age squared	0.001	0.001	0.001	
	(0.001)	(0.001)	(0.001)	
Male	0.037***	0.042*	0.038**	0.038**
	(0.004)	(0.017)	(0.015)	(0.015)
College	-0.000	-0.000	-0.005	-0.005
	(0.004)	(0.004)	(0.005)	(0.005)
Married	-0.018***	-0.018***	-0.018***	-0.018***
	(0.005)	(0.005)	(0.005)	(0.005)
Full age dummies	No	No	No	Yes
Income decile dummies	No	No	Yes	Yes
First stage F-stat		41.4	42.6	42.6
Individuals	10,337	10,337	9,924	9,924
Observations	59,566	59,566	57,418	57,418

Estimated coefficients from regressions of a dummy for hospitalization onto retirement.
Column (1) contains estimates from model (3) without use of instrument (eq. (4) is not used) Column (2)-(4) contain estimates from model (3) instrumenting retirement with a dummy for (birthday $\geq$ July 1st 1939) (eq. (4) as the first stage). Estimated for the 1939 cohort, age 65 to 70. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.4: Reform strategy: GP visits

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
Retired	0.772***	-0.958	-0.883	-0.883
	(0.086)	(2.002)	(2.000)	(2.000)
Age	-6.312***	-3.645	-3.952	
	(1.385)	(3.462)	(3.392)	
Age squared	0.049***	0.030	0.032	
	(0.010)	(0.025)	(0.024)	
Male	0.019	-0.347	-0.158	-0.158
	(0.088)	(0.431)	(0.348)	(0.348)
College	-0.427***	-0.440***	-0.328**	-0.329**
	(0.089)	(0.092)	(0.109)	(0.109)
Married	-0.109	-0.084	-0.060	-0.060
	(0.099)	(0.104)	(0.111)	(0.111)
Full age dummies	No	No	No	Yes
Income decile dummies	No	No	Yes	Yes
First stage F-stat		32.1	33.3	33.3
Individuals	10,088	10,088	9,709	9,709
Observations	47,497	47,497	45,865	45,865

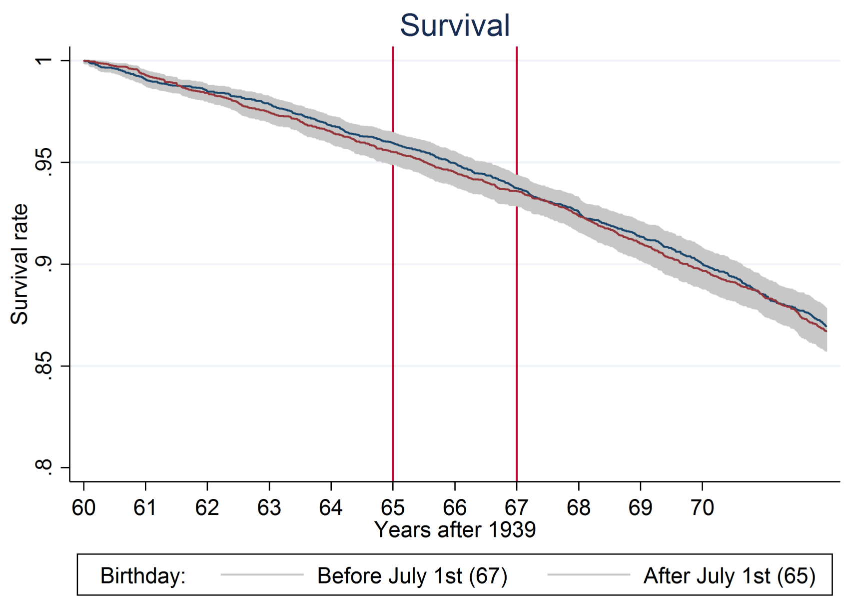
Estimated coefficients from regressions of the number of GP visits onto retirement.
Column (1) contains estimates from model (3) without use of instrument (eq. (4) is not used) Column (2)-(4) contain estimates from model (3) instrumenting retirement with a dummy for (birthday $\geq$ July 1st 1939) (eq. (4) as the first stage). Estimated for the 1939 cohort, age 65 to 70. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.5: Reform strategy: mortality

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
Retired	0.006***	-0.009	-0.008	-0.008
	(0.001)	(0.020)	(0.020)	(0.020)
Age	-0.004	0.017	0.015	
	(0.026)	(0.036)	(0.036)	
Age squared	0.000	-0.000	-0.000	
	(0.000)	(0.000)	(0.000)	
Male	0.008***	0.005	0.008*	0.008*
	(0.001)	(0.004)	(0.003)	(0.003)
College	-0.005***	-0.005***	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
Married	-0.007***	-0.007***	-0.007***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Full age dummies	No	No	No	Yes
Income decile dummies	No	No	Yes	Yes
First stage F-stat		41.3	42.5	42.4
Individuals	10,313	10,313	9,908	9,908
Observations	59,453	59,453	57,334	57,334

Estimated coefficients from regressions of the 1-year mortality rate onto retirement.
Column (1) contains estimates from model (3) without use of instrument (eq. (4) is not used) Column (2)-(4) contain estimates from model (3) instrumenting retirement with a dummy for (birthday $\geq$ July 1st 1939) (eq. (4) as the first stage). Estimated for the 1939 cohort, age 65 to 70. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure B.1: Reform strategy: survival by birthday threshold



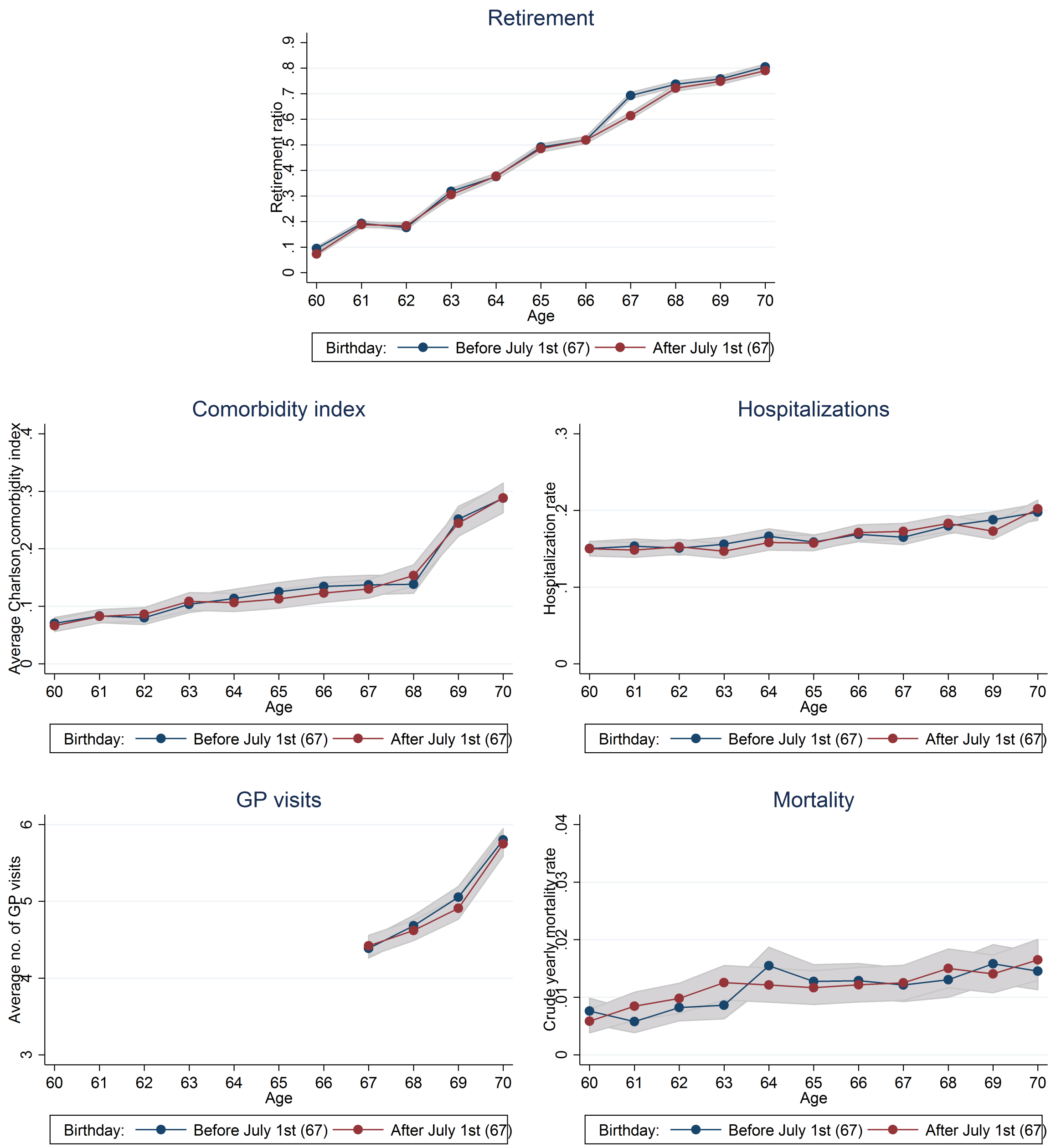
Survival ratio conditional on surviving until December 31st 1999 split by birthday threshold with 95% confidence intervals as shaded area. Vertical lines depict ages when the 1999 retirement reform induces differences in retirement takeup. Sample: 1939 cohort, individuals eligible for ERP excluded, individuals receiving disability pension before age 60 excluded.

Table B.6: Reform strategy: placebo regressions

	(1) Charlson	(2) Hospitalized	(3) GP visits	(4) Mortality
Retired	0.073	-0.082	1.104	-0.008
	(0.370)	(0.177)	(2.519)	(0.041)
First stage F-stat	10.7	10.7	19.8	10.8
Individuals	10,570	10,570	10,202	10,563
Observations	61,238	61,238	38,804	61,175

Each column contains estimates from 2SLS regressions of the dependent variable onto a dummy for retirement and controls, instrumenting retirement with the 1938 placebo threshold. Estimated for the 1938 cohort, age 65 to 70. Controls: Full set of age and income decile dummies, dummy for male, college and married. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

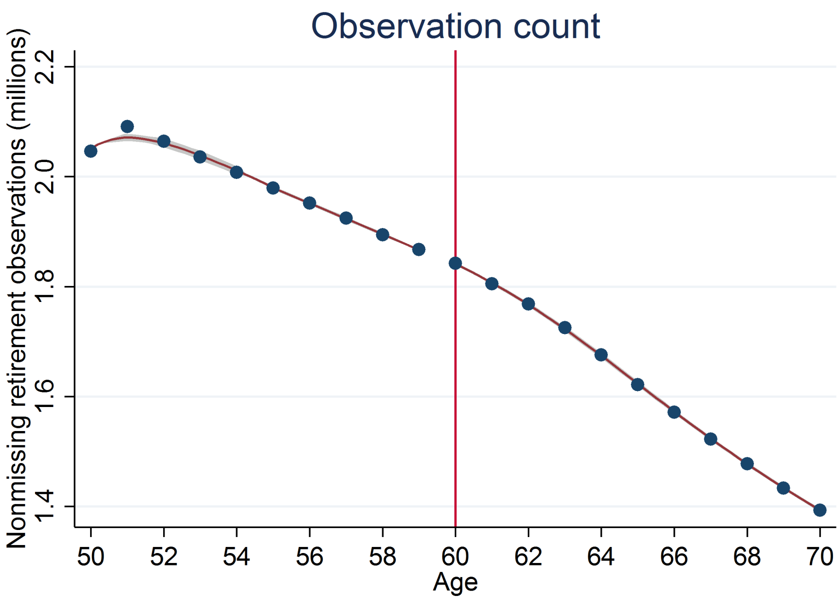
Figure B.2: Reform strategy: placebo cohort by birthday threshold



Placebo exercise for the 1938 cohort. Mean values of the dependent variable conditional on age and birthday in 1938 with 95% confidence intervals as shaded area. Sample: 1938 cohort, individuals eligible for ERP excluded, individuals receiving disability pension before age 60 excluded, individuals excluded in the year of death.

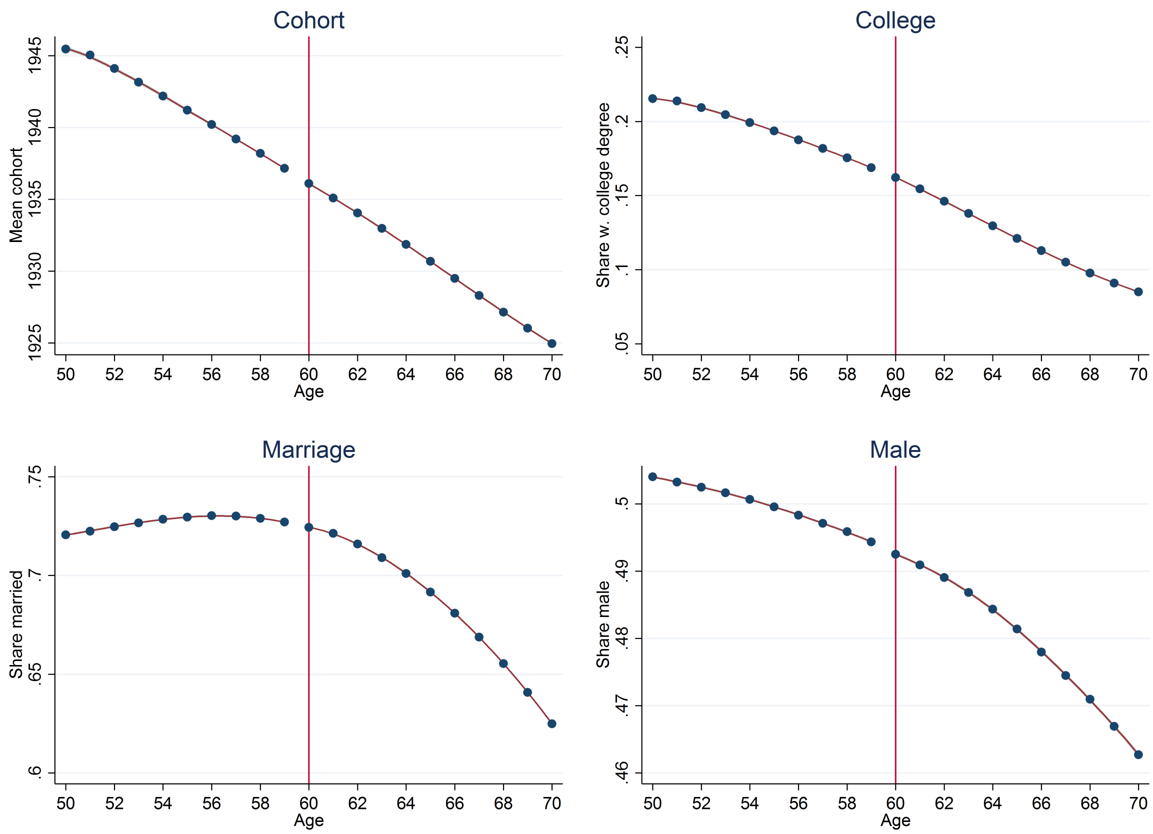
Appendix C: RD robustness

Figure C.1: RD strategy: observation count by age



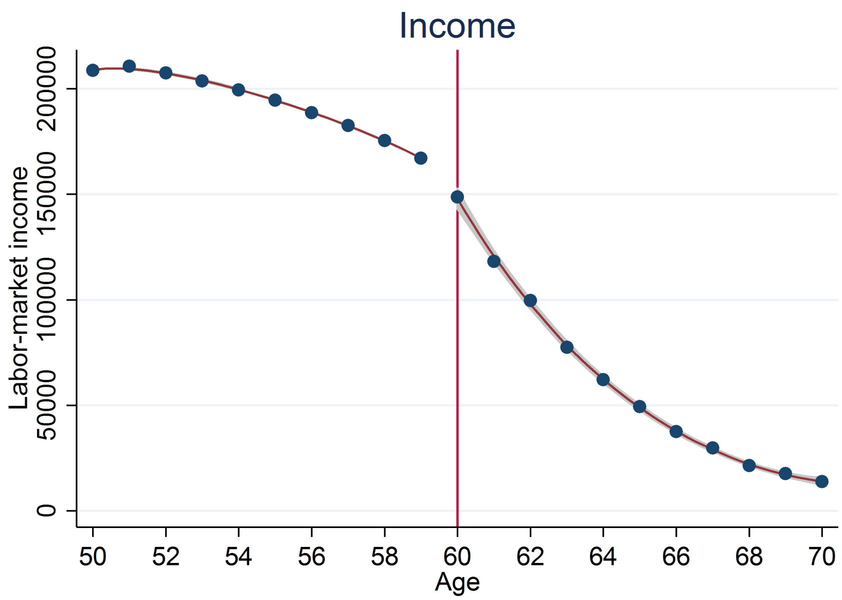
Number of non-missing retirement observations by age in millions. Lines show the local quadratic fit using a triangular kernel and a bandwidth of 5 years with 95% confidence intervals as shaded area. Sample: full/RD sample.

Figure C.2: RD strategy: covariates by age



Mean values of non-health variables by age. Lines show the local quadratic fit using a triangular kernel and a bandwidth of 5 years with 95% confidence intervals as shaded area. Sample: full/RD sample.

Figure C.3: RD strategy: labor market income by age



Mean labor market income by age. Lines show the local quadratic fit using a triangular kernel and a bandwidth of 5 years with 95% confidence intervals as shaded area. Sample: full/RD sample.

Table C.1: RD strategy: covariate tests

	(1) Income	(2) College	(3) Married	(4) Male	(5) Cohort
Retired	-54,384***	0.003**	0.000	-0.001	0.000
	(1,719)	(0.001)	(0.001)	(0.001)	(0.032)
First stage F-stat	223,542	215,928	215,928	215,928	215,928
Individuals	2,385,139	2,410,693	2,410,693	2,410,693	2,410,693
Observations	16,247,715	16,453,333	16,453,333	16,453,333	16,453,333

Estimated coefficients on retirement from local 2SLS regressions of the dependent variable onto a dummy for retirement and a quadratic polynomial in age on either side of the threshold (age 60), instrumenting retirement with the age threshold (age $\geq$ 60). Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.2: RD strategy: covariate tests, balanced sample

	(1) Income	(2) College	(3) Married	(4) Male	(5) iv_cohort
Retired	-46,004***	-0.000	0.001	-0.000	0.000
	(1,376)	(0.000)	(0.001)	(.)	(.)
First stage F-stat	191,836	191,835	191,835	191,835	191,835
Individuals	1,094,905	1,094,906	1,094,906	1,094,906	1,094,906
Observations	9,854,063	9,854,154	9,854,154	9,854,154	9,854,154

Estimated coefficients on retirement from local 2SLS regressions of the dependent variable onto a dummy for retirement and a quadratic polynomial in age on either side of the threshold (age 60), instrumenting retirement with the age threshold (age $\geq$ 60). Sample: balanced sample of individuals observed at all ages from 55 to 65. Standard errors clustered on the individual level in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.3: RD strategy: Charlson index

Charlson Index	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS	FE-2SLS
Kernel	Triangular	Triangular	Triangular	Rectangular	Triangular	Triangular	Triangular
Bandwidth	5	5	10	5	5	5	5
Retired	0.089***	-0.002	-0.003	-0.002	0.001	-0.001	-0.015**
	(0.001)	(0.005)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)
Age (η_1)	-0.009***	0.006***	0.006***	0.006***	0.006***	0.006***	0.018***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
Age squared (η_2)	-0.003***	0.000	0.000***	0.000		0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Age, right (η_3)	0.007***	0.002*	0.002***	0.002***	0.002***	0.001	0.000
	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
Age squared, right (η_4)	0.004***	-0.000	-0.000***	-0.000		0.000	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Additional controls	No	No	No	No	No	Yes	No
First stage F-stat		215,928	302,205	242,075	332,103	218,167	224,280
Individuals	2,410,693	2,410,693	3,099,474	2,548,875	2,410,693	1,890,020	2,263,614
Observations	16,453,333	16,453,333	34,257,679	20,053,713	16,453,333	13,902,834	16,306,254

Estimated coefficients from local regressions of the Charlson index onto retirement.
Column (1) contains OLS estimates from model (5) with age specification (7) without use of instruments eq. (6) is not used).
Column (2)-(4) contain estimates from model (5) with age specification (7) instrumenting retirement with the threshold dummy eq. (6) is the first stage. Models vary according to bandwidth and kernel
Column (5) is similar to column (2) but with a linear age specification. Column (6) is similar to column (2) but controlling for full sets of year and income decile dummies and dummies for male, college and married. Column (7) is similar to column (2) but adding individual specific FE's. Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.4: RD strategy: hospitalization

Hospitalized	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS	FE-2SLS
Kernel	Triangular	Triangular	Triangular	Rectangular	Triangular	Triangular	Triangular
Bandwidth	5	5	10	5	5	5	5
Retired	0.067***	-0.011**	-0.013***	-0.013***	-0.012***	-0.012**	-0.010*
	(0.000)	(0.004)	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)
Age (η_1)	-0.012***	0.002**	0.002***	0.002***	0.002***	0.002**	0.008***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Age squared (η_2)	-0.003***	-0.000	0.000***	0.000		0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Age, right (η_3)	0.005***	0.001	0.001***	0.001**	0.002***	0.001	0.001
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Age squared, right (η_4)	0.004***	0.000	-0.000	0.000		0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
Additional controls	No	No	No	No	No	Yes	No
First stage F-stat		215,928	302,205	242,075	332,103	218,167	224,280
Individuals	2,410,693	2,410,693	3,099,474	2,548,875	2,410,693	1,890,020	2,263,614
Observations	16,453,333	16,453,333	34,257,679	20,053,713	16,453,333	13,902,834	16,306,254

Estimated coefficients from local regressions of a dummy for hospitalization onto retirement.
Column (1) contains OLS estimates from model (5) with age specification (7) without use of instruments eq. (6) is not used).
Column (2)-(4) contain estimates from model (5) with age specification (7) instrumenting retirement with the threshold dummy eq. (6) is the first stage. Models vary according to bandwidth and kernel
Column (5) is similar to column (2) but with a linear age specification. Column (6) is similar to column (2) but controlling for full sets of year and income decile dummies and dummies for male, college and married. Column (7) is similar to column (2) but adding individual specific FE's. Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.5: RD strategy: GP visits

GP visits	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	2SLS	2SLS	2SLS	2SLS	2SLS	FE-2SLS
Kernel	Triangular	Triangular	Triangular	Rectangular	Triangular	Triangular	Triangular
Bandwidth	5	5	10	5	5	5	5
Retired	1.678***	-0.348***	-0.313***	-0.258**	-0.307***	-0.371***	-0.470***
	(0.013)	(0.094)	(0.059)	(0.079)	(0.052)	(0.094)	(0.088)
Age (η_1)	-0.247***	0.080***	0.059***	0.056***	0.059***	0.078***	0.191***
	(0.008)	(0.016)	(0.006)	(0.011)	(0.004)	(0.016)	(0.014)
Age squared (η_2)	-0.058***	0.005	-0.000	-0.001		0.004	0.012***
	(0.002)	(0.003)	(0.001)	(0.002)		(0.003)	(0.003)
Age, right (η_3)	0.137***	0.060***	0.106***	0.090***	0.110***	0.057***	0.056***
	(0.012)	(0.012)	(0.007)	(0.009)	(0.005)	(0.012)	(0.011)
Age squared, right (η_4)	0.083***	0.005	0.002*	0.006*		0.005	0.001
	(0.002)	(0.004)	(0.001)	(0.002)		(0.004)	(0.004)
Additional controls	No	No	No	No	No	Yes	No
First stage F-stat		33,478	45,725	37,658	57,549	37,768	41,507
Individuals	987,030	987,030	1,606,520	1,114,774	987,030	978,747	842,685
Observations	3,745,435	3,745,435	7,401,122	4,517,631	3,745,435	3,721,957	3,601,090

Estimated coefficients from local regressions of the number of GP visits onto retirement.
Column (1) contains OLS estimates from model (5) with age specification (7) without use of instruments eq. (6) is not used).
Column (2)-(4) contain estimates from model (5) with age specification (7) instrumenting retirement with the threshold dummy eq. (6) is the first stage. Models vary according to bandwidth and kernel
Column (5) is similar to column (2) but with a linear age specification. Column (6) is similar to column (2) but controlling for full sets of year and income decile dummies and dummies for male, college and married. Column (7) is similar to column (2) but adding individual specific FE's. Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.6: RD strategy: mortality

1-year mortality	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
Kernel	Triangular	Triangular	Triangular	Rectangular	Triangular	Triangular
Bandwidth	5	5	10	5	5	5
Retired	0.013***	0.001	0.001	0.001	0.001	0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age (η_1)	-0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age squared (η_2)	-0.000***	0.000	0.000***	0.000*		0.000
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)
Age, right (η_3)	0.001***	-0.000	-0.000	-0.000	0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age squared, right (η_4)	0.001***	0.000	0.000***	0.000		0.000
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)
Additional controls	No	No	No	No	No	Yes
First stage F-stat		223,537	308,304	248,708	338,616	230,949
Individuals	2,385,218	2,385,218	3,067,203	2,522,426	2,385,218	1,874,031
Observations	16,248,227	16,248,227	33,806,799	19,801,716	16,248,227	13,742,568

Estimated coefficients from local regressions of 1-year mortality index onto retirement.
Column (1) contains OLS estimates from model (5) with age specification (7) without use of instruments eq. (6) is not used).
Column (2)-(4) contain estimates from model (5) with age specification (7) instrumenting retirement with the threshold dummy eq. (6) is the first stage. Models vary according to bandwidth and kernel
Column (5) is similar to column (2) but with a linear age specification. Column (6) is similar to column (2) but controlling for full sets of year and income decile dummies and dummies for male, college and married. Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.7: RD strategy: placebo threshold

	Charlson		Hospitalization		GP visits		Mortality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age threshold	60	57	60	57	60	57	60	57
Retired	-0.000	-0.608	-0.011***	0.640	-0.309***	26.211	0.001	0.313
	(0.004)	(1.599)	(0.003)	(1.379)	(0.074)	(17.573)	(0.001)	(0.377)
First stage F-stat	269,345	5.3	269,345	5.3	43,920	3.7	277,847	5.8
Individuals	2,129,660	2,220,773	2,129,660	2,220,773	721,638	719,398	2,105,902	2,200,147
Observations	9,176,476	9,616,440	9,176,476	9,616,440	2,125,924	2,051,027	9,063,237	9,520,737

Placebo exercise for the RD-design estimates. Each column contains estimates from model (5) with a linear age specification on either side of the threshold ($\eta_2 = \eta_4 = 0$ in equation (7)) instrumenting retirement with a threshold dummy. Columns vary by dependent variable and cutoff age. All models estimated with a 2-year bandwidth and rectangular kernel.
For each dependent variable, the left/odd column contains the estimates using the (true) age 60 cutoff. These estimates should be qualitatively similar to the main 2SLS estimates in table 3. The right/even columns contains the estimates from a similar model using a placebo cutoff, age 57. Sample: full/RD sample. Standard errors clustered on the individual level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix D: Details on health measures

GP visits Each procedure in the primary healthcare sector has a unique 6-digit number. The first two digits describe the type of doctor or specialist. The identification number for general practitioners is 80. The subsequent 4 digits describe the type of procedure. I count visits as consultations with the GP, not including email or telephone consultations, using the procedure variable "SPECIALE" from Statistics Denmark. Specifically, I include the following procedure codes: 800100, 800101, 800102, 800103, 800104, 800106, 800107, 800108 which are all consultations at the GP's office.

Charlson Index The Charlson index weighs serious diagnoses. I identify these diagnoses as described in the main text and weigh them accordingly. The diagnoses are: Myocardial infarction, Congestive heart failure, Peripheral vascular disease, Cerebrovascular disease, Dementia, Chronic pulmonary disease, Connective tissue disease, Ulcer disease, Mild liver disease, Diabetes type 1 and 2, Hemiplegia, Moderate to severe renal disease, Diabetes with end organ damage type 1 and 2, Any tumor, Leukemia, Lymphoma, Moderate to severe liver disease, Metastatic solid tumor, AIDS. The codes and weights for each type of diagnosis are presented in table C.1 below.

Table C.1: Charlson ICD codes

Charlson group	ω_d	ICD8	ICD10
Myocardial Infarction	1	410	I21;I22;I23
Congestive heart failure	1	42709-42711; 42719;42899;78249	I50;I110;I130; I132
Peripheral Vascular Disease	1	440-445;	I70-I74; I77
Cerebrovascular Disease	1	430-438;	I60-I69; G45;G46
Dementia	1	29009-29019; 29309	F00-F03;F051;G30
Chronic Pulmonary Disease	1	490-493;515-518	J40-J47;J60-J67; J684;J701;J703;J841; J920;J961;J982;J983
Connective Tissue Disease	1	712;716;734;446; 13599	M05;M06;M08;M09; M30-M36;D86
Ulcer Disease	1	531-534;53091; 53098	K25-K28;K221
Mild Liver Disease	1	571;57301;57304	B18;K71;K73;K74; K700-K703;K709;K760
Diabetes type I or II	1	24900;24906;24907; 24909; 25000; 25006;25007;25009	E100;E101; E109;E110; E111;E119
Hemiplegia	2	344	G81;G82
Renal Disease (moderate to severe)	2	403;404;580-584; 59009;59319;75310; 75311;75312-75319; 792	I12;I13;N00;N01-N05; N07;N11;N14;N17-N19; Q61
Diabetes with organ damage	2	24901-24905;24908; 25001-25005;25008	E102-E108; E112-E118
Cancer, Any tumor	2	140-I94	C00-C75
Cancer, Leukemia	2	204-207	C91-C95
Cancer, Lymphoma	2	200-203;27559	C81-C85;C88;C90;C96
Liver Disease (moderate to severe)	3	7000;7002;7004;7006; 7008;45600-45604; 45605-45609;57300	B150;B160;B162;B190; K704;K766;I85;K72
Cancer, Metastatic Solid Tumor	6	195-199	C76-C80
AIDS	6	7983	B21-B24

Implementation follows Johansen and Fynbo (2011)