Discussion Paper Series

IZA DP No. 18390

February 2026

Delayed Retirement: Effects on Health and Healthcare Utilization

Anne Katrine Borgbjer
Aarhus University

Hans Schytte Sigaard
Aarhus University

Michael Svarer
Aarhus University and IZA@LISER

Rune Majlund Vejlin
Aarhus University and IZA@LISER

IZA NETWORK @ LISER
ISSN: 2365-9793

Delayed Retirement: Effects on Health and Healthcare Utilization*

Abstract
This paper estimates the effect of a reform-induced increase in the early retirement age (ERA) on labor supply, health, and healthcare utilization using detailed Danish administrative data and a regression discontinuity design. We show that while raising the ERA successfully increased employment, it also led to spillovers into other public transfers and increased the number of self-supporting individuals. We find that the increased ERA led to small and insignificant effects on GP visits and the use of painkillers, as well as borderline significant, small positive effects on the use of antidepressants and CVD medicine. Further analysis shows that individuals who were employed due to the reform had lower pre-reform income and wealth, while the individuals who were not employed despite being affected by the reform were characterized by worse health before the reform announcement. We argue that possibilities for exiting employment serve as a potentially important mitigating mechanism for health and healthcare utilization effects by sorting vulnerable individuals out of employment.

JEL classification
I18, J18, J26

Keywords
retirement reforms, health, healthcare utilization

Corresponding author
Rune Majlund Vejlin
rvejlin@econ.au.dk

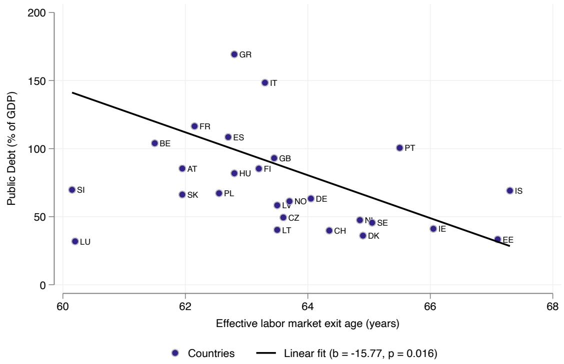
* We thank Marianne Simonsen, Torben M. Andersen, Maja Weemes Grøtting, and Lisa Laun, as well as participants at DGPE 2023, Aarhus University, Pensjonsforum, 16th Nordic Summer Institute in Labor Economics, 15th Nordic Meeting on Register Data and Economic Modelling, ESPE 2024, EALE 2024, Harvard University, Nordic Public Policy Symposium 2024, and SOFI, for valuable comments and suggestions. We also thank the ECONAU project database (Department of Economics and Business Economics, Aarhus University) for making their data available to us. Anne Katrine Borgbjerg gratefully acknowledges financial support from the Rockwool Foundation.

1 Introduction

Aging populations pose a challenge to the financial sustainability of public pension systems in most developed countries. In response, many countries have begun implementing retirement reforms aimed at increasing employment among the elderly, for example, by raising the age of eligibility for publicly provided pension schemes (OECD, 2023). Figure 1 illustrates both of these points. Figure 1a illustrates the relationship between public debt and effective retirement ages across all European OECD countries. A clear negative association emerges between retirement age and the level of public debt.[1] Figure 1b presents the current and planned retirement ages for the same group of countries. Retirement ages vary by roughly five years, and many countries have enacted or announced plans to raise them.

Figure 1: Public Debt and Retirement Ages

(a) Public Debt and Effective Retirement Ages



(b) Actual and Planned Retirement Ages

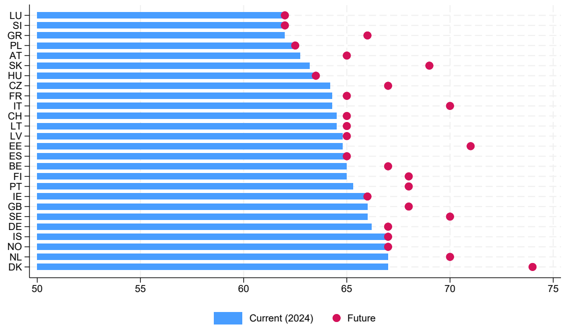


Figure 1a shows the population-weighted relationship between public debt and the effective retirement age for all European OECD countries. Data is from the OECD (2025). Figure 1b shows the actual and planned retirement ages. Planned retirement ages are calculated for a person aged 22. Data is from OECD (2024).

However, pension schemes exist fundamentally to allow individuals to withdraw from a physically or mentally demanding labor market as health deteriorates with age (Blundell et al., 2016). Retirement reforms that extend the effective working age are therefore not without potential health consequences, but despite their widespread adoption, the health-related implications remain a subject of debate with no clear consensus on their extent or direction. A critical policy challenge lies in designing retirement reforms that not only alleviate the financial pressure on social security systems but also address potential adverse health effects on the elderly.

In this paper, we examine the labor supply impacts, the presence of health and healthcare utilization effects, and the underlying mechanisms of increasing the eligibility age for claiming early retirement benefits in Denmark. Denmark provides an ideal setting for this investigation, since it started to imple-

[1] Note, the US would be an outlier in this graph with both a very high level of debt and a high retirement age.

ment retirement reforms very early and also has population-wide register data on labor market activities, public transfers, and, most importantly, health and healthcare utilization measures. In particular, we use a reform implemented in 2014 that introduced sharp discontinuities in early retirement ages based on birth dates.

Despite the health and healthcare utilization effects of increasing the retirement age being a key policy concern, there remains a lack of consensus regarding the magnitude and direction of these effects in the existing literature. On one hand, Grip et al. (2012), Shai (2018), Carrino et al. (2020), Barschkett et al. (2022), and Serrano-Alarcón et al. (2023) show that an increased retirement age can lead to deterioration in overall health status, worsened mental health, increased healthcare spending, and more injuries.[2] Conversely, Hernaes et al. (2013), Bertoni et al. (2018), Hagen (2018), Bozio et al. (2021), Bauer and Eichenberger (2021), and Ci (2022) find that increases in the retirement age can reduce obesity and BMI, increase self-reported satisfaction with health, reduce prescription drug usage, and have no effect on mortality.[3] These differing effects might be due to contextual factors, such as differing social security systems (e.g., the availability of alternative exit routes from employment), the roll-out of the reform (whether it was gradually phased in, allowing individuals to adjust labor market behavior, savings, and health investments, or introduced abruptly), or the population affected by the reform (e.g., male construction workers in Bauer and Eichenberger (2021) versus female local government workers in Hagen (2018)). We provide an overview of all studies cited above in Appendix A, including information on contextual factors.[4]

Similarly, while Atalay et al. (2019), Cremers et al. (2024), and Banks et al. (2025) find a negative effect of retirement on cognition and mental health, Hallberg et al. (2015), Bíró and Elek (2018), Nielsen (2019), Frimmel and Pruckner (2020), and Kuusi et al. (2020) find that retirement leads to a decrease in in- and outpatient care, and reduces pharmaceutical expenditures, the use of antidepressants, doctor visits and hospitalizations.[5]

[2] Specifically, Grip et al. (2012) show that a reduction in pension rights from a reform in the Dutch pension system led to a strong deterioration in mental health for the affected workers, and Shai (2018) shows that an increase in the retirement age for men in Israel led to a deterioration in health status and increased healthcare spending. Carrino et al. (2020) show that an increase in the state pension age in the UK led to higher risk of depression, Barschkett et al. (2022) find that a higher retirement age increased obesity, musculoskeletal diseases, and mental health diagnoses for German women, and Serrano-Alarcón et al. (2023) show that a higher normal retirement age in Italy increased injuries and mental health hospitalizations for women.

[3] Specifically, Hernaes et al. (2013), Hagen (2018), and Bozio et al. (2021) find no effect of changing the retirement age on mortality, Bertoni et al. (2018) find an increase in the minimum retirement age in Italy reduced obesity and increased self-reported satisfaction with health, Bauer and Eichenberger (2021) show that among Swiss construction workers, a decrease in the retirement age worsened self-reported health, implying under symmetry that a higher retirement age improves health, and Ci (2022) finds that an increase in the retirement age in the US lead to a reduction in prescription drug usage, increased subjective well-being, and reduced BMI.

[4] Additionally, Brugiavini et al. (2026) do not find evidence that health inequality in Italy increased from 2004 to 2022, a period where pensions reforms increased retirement ages and decreased benefit generosity.

[5] Specifically, Atalay et al. (2019) find a negative effect of retirement on cognition of Australians while Cremers et al. (2024)

In this paper, we utilize Danish administrative register data, employing a regression discontinuity (RD) design, to exploit a reform that discontinuously increased the early retirement age (ERA) based on individuals' date of birth. The 2006 Welfare Agreement led to planned increases in retirement ages beginning from 2019; however, the 2011 Retirement Reform was announced in 2011 and progressively increased the ERA by ½ year annually, already from 2014 to 2019, from a level of 60 years, introducing six-month discontinuous changes in the ERA for cohorts born on January 1, 1954, and after. Note that although the individuals affected are, by definition, younger than the statutory retirement age in Denmark, they are approximately the same age as most Europeans at their actual retirement (see Figure 1b). This supports the external validity of our estimates. Our analysis focuses on individuals unaffected by the 2006 Welfare Agreement but impacted by the 2011 Retirement Reform. We use the reform-induced anticipated increase in the ERA by comparing health and healthcare utilization outcomes for individuals born less than two months before and after a July 1 cutoff, thereby comparing individuals who are similar in every aspect except for the sharp difference in ERA. We examine a large set of health and healthcare utilization outcomes: general practitioner (GP) visits, use of painkillers, antidepressants, and cardiovascular disease (CVD) medicine, with the prescription drug outcomes measured by defined daily doses (DDDs).[6] By applying the RD strategy to detailed Danish administrative register data on labor market outcomes, background characteristics, and healthcare utilization, we divide our findings into three parts.

First, we show that the ½ year increase in the ERA increased employment by 17 percentage points (27%) in the half-year interval after the extended ERA,[7] but also led to an uptake of other public transfers of 4 percentage points (41%). The majority of the increase in other transfers is caused by an uptake in ordinary transfers (unemployment insurance or cash benefits), rather than an uptake in health transfers (such as sickness benefits or disability pensions). This is not unexpected, given the greater accessibility of ordinary transfers compared to health transfers. We also find that the ½ year increase in the ERA

find that retirement increases prescription medications for hypertension and mental health, but decreases the use of painkillers in Denmark. Banks et al. (2025) study the same UK pension reform as Carrino et al. (2020), but they find no adverse effects of the reform on mental health. Instead, they report that working longer improves cognitive outcomes and physical mobility. Hallberg et al. (2015) find that an early retirement offer for military officers in Sweden decreased mortality and inpatient care, Bíró and Elek (2018) find that for women in Hungary, retirement leads to a decrease in outpatient care, inpatient care, and pharmaceutical expenditures, Nielsen (2019) finds that early retirement leads to decreases in general practitioner (GP) visits and hospitalization in Denmark, Frimmel and Pruckner (2020) find that early retirement in Austria leads to a reduction in doctor visits, hospitalization, and expenditure for outpatient medical attendance, and Kuusi et al. (2020) find that retirement in Finland moderately decreases the use of antidepressants.

[6] In Appendix D.1, we present the effects of increased ERA on additional health and healthcare utilization outcomes: binary indicators for the use of painkillers, antidepressants, and CVD medicine, as well as the Charlson Comorbidity Index (CCI), stroke incidence, and mortality.

[7] This increase in employment is similar to the findings of Danish Economic Councils (2021) who analyzed the 2011 reform using a similar RD strategy.

increased the share of individuals who are self-supporting (i.e., individuals who are neither employed nor receiving public transfers) by 2 percentage points (26%).

Second, our findings indicate that the increase in ERA had no effect on GP visits or usage of painkillers, but led to borderline significant, small increases in the usage of antidepressants and CVD medicine. Our dynamic effects show that the effects on antidepressants and CVD medicine begin around the ERA increase, and while the effect on antidepressants vanishes after 4.5 years, the effect on CVD medicine appears more persistent. We find evidence of treatment effect heterogeneity estimated using a causal forest. Merging our data with O*NET data, we find, for instance, that workers in occupations characterized by a high degree of routine and low decision-making freedom increase their use of antidepressants more in response to delayed retirement. Treatment effect heterogeneity is important because different retirement reforms target different populations; thus, the lack of consensus across studies may reflect not only methodological and institutional differences but also heterogeneous treatment effects arising from different populations being affected.

Third, we highlight important contextual factors of the reform using a complier analysis as suggested by Marbach and Hangartner (2020). We define compliers as the group who are employed in the half-year interval after the ERA increase *because* of the increased ERA, while, regardless of the reform, always-takers are employed, and never-takers are not.[8] We show that the complier group is characterized by having a larger share of women, lower education, and lower disposable income and net wealth before the reform announcement. Differences in income and net wealth suggest that the reform has a greater impact on financially vulnerable individuals. Moreover, the never-takers of the reform, i.e., those who are not employed regardless of the reform, are characterized by worse health and healthcare utilization outcomes prior to the reform announcement. In general, compliers are in better health than never-takers, but in worse health than always-takers.

The characteristics of compliers represent important contextual factors when designing future retirement reforms. If individuals in worse health are more negatively impacted by increases in the retirement age, implementing "stricter" retirement reforms, i.e., reforms that induce more individuals to extend employment (e.g., by changing the level of pensions), could have more pronounced effects on health and healthcare utilization. We examine this point by employing the marginal treatment effect (MTE) framework in a local instrumental variable (LIV) approach to show that controlling for selection into employment can uncover important heterogeneity. The MTE results suggest that the individuals who

[8]While we cannot observe compliers at the individual level, we can calculate complier covariate means by subtracting the weighted covariate means for always-takers and never-takers from the covariate mean for the entire sample, assuming independence of the instrument and monotonicity.

avoided employment are the ones who would have experienced the largest effects on antidepressants and CVD medicine.

We contribute to the literature studying the health effects of retirement reforms in several ways. First, utilizing high-quality, detailed administrative data covering the entire Danish population, we examine a broad set of health and healthcare utilization outcomes applying the latest advancements in the RD literature. The administrative data, with daily granularity for prescriptions and hospital diagnoses and weekly for GP visits, surpass the limitations of prior studies relying on annual data. Notably, we calculate *exact* age-specific effects of increased retirement ages. Exact age-specific effects overcome limitations of the annual data used in previous studies, as annual data may obscure differences, for example, between two individuals turning 60 in the same year but several months apart. Such temporal discrepancies are crucial, particularly for individuals approaching retirement, a period where health may increasingly deteriorate. Additionally, the granularity of data also enables using the local randomization framework for RD designs, which yields plausible estimates of the effects of increased retirement ages with high precision. Using administrative data also overcomes problems with self-assessed health measures, low response rates, and lack of follow-up, which are prevalent in survey data.

Second, previous literature often focus on sudden, drastic reforms (such as, e.g., Grip et al. (2012), Shai (2018), and Serrano-Alarcón et al. (2023)) which are desirable from an econometric perspective as they provide large exogenous variation, but the results may not extrapolate to long-run effects or to reforms that are gradually implemented over long time horizons which are more commonly implemented and more desirable from a policy perspective. In both settings, individuals will have time to adjust their behavior. We use a reform that is announced well ahead of time and gradually implemented, but with six-month discontinuities that allow credible identification of health and healthcare utilization effects. Initially, the three-year announcement period may not seem like a positive attribute for empirical research design. However, we argue that the slow rollout has two advantages. First, it more closely resembles how reforms are typically announced, making our results easier to extrapolate to other reforms. Second, and more importantly, the resulting estimates are likely closer to long-run effects, since individuals can adjust their behavior, e.g., savings, health investments, and labor market behavior, in the long run, taking the reform into account. Typically, the long-run effect of retirement reforms are of main policy interest, and not the reform-induced effect per se.

Third, While the effects of increasing retirement ages on employment and spillovers on other public transfer programs is well-documented (see e.g., Mastrobuoni (2009) and Manoli and Weber (2016a) for employment effects and Duggan et al. (2007), Vestad (2013), and Staubli and Zweimüller (2013) for spillovers), we show that the never-takers of a retirement reform are characterized by worse pre-reform health and healthcare utilization outcomes and to highlight the potential mitigating role of the possibility to exit employment, either by spillovers into other transfer programs or the ability to be self-supporting.[9] This potential mitigating mechanism can be crucial for comparing studies from countries that differ in their social security systems or mandated savings, as it suggests countries with smaller opportunities to transition into other transfers or to self-finance retirement will exhibit larger effects on health and healthcare utilization of increased retirement ages.

The Danish 2011 retirement reform has previously been used to study the effects of increasing the retirement age on employment and public transfers (Athey et al., 2020, Danish Economic Councils, 2021), joint retirement decisions (García-Miralles and Leganza, 2024a), and pension contributions (García-Miralles and Leganza, 2024b). The study most closely related to ours, Cremers et al. (2024), also investigates health outcomes using the 2011 reform. However, while they apply a four-year window around the January 1 cutoff, we focus on a much narrower 50-day window using the July 1 cutoff, providing much clearer identification. Our contribution relative to this existing literature is to leverage recent methodological advances in econometrics to bridge the labor supply and health literatures on the effects of increasing retirement ages.

Our paper is structured as follows. In Section 2, we introduce the Danish institutional context and the data used, Section 3 outlines our empirical strategy, and in Section 4, we present our results. We discuss mechanisms in Section 5, before concluding in Section 6.

2 Institutional Setting and Data

2.1 The Danish Pension System

The Danish retirement system consists of three pillars: 1. Public pensions (including early retirement and supplements), 2. Occupational pensions, and 3. Private pensions and other savings. Each of the three pillars is covered in turn below, the first pillar being of most relevance for this paper and receiving the most emphasis accordingly.

Public pensions are defined benefit (DB) pensions financed through taxes, available to individuals who have reached the statutory retirement age (SRA), which was 65 years in 2011. Public pensions

[9] Our findings complement those of Ollonqvist et al. (2025) that study the relationship between health status and financial incentives to delay retirement. They find that most individuals, including those in poor health, generally respond to financial incentives to delay retirement, although individuals with prolonged sickness absence are less responsive. While Ollonqvist et al. (2025) exploits individual-level variation in the change in pension wealth from postponing retirement by one year (generated by a multifaceted Finnish reform that simultaneously changed retirement ages, accrual rates, and other key pension parameters), our study leverages a sharp discontinuous increase in the early retirement eligibility age by date of birth.

provide a flat universal rate and means-tested supplements. The largest supplementary scheme is the savings-based Market Supplementary Pension Fund (*Arbejdsmarkedets Tillægspension*, ATP), which is mandatory by law for almost all wage earners. ATP is funded by individual contributions while active on the labor market. It is usually paid 1/3 by the employee and 2/3 by the employer. Public pension, ATP, and additional supplements are all established by the government to lower poverty and ensure risk sharing among Danish citizens.

Early retirement (also called "Post Employment Wage", PEW) is a voluntary scheme introduced in 1979 to provide workers the opportunity to retire early. The primary requirements for early retirement are being a member of an unemployment insurance fund, having paid contributions for at least 30 years, and meeting the early retirement age (60 years in 2011). The scheme gained more traction than intended, and reforms in 1999, 2006, and 2011 all had a focus on reducing the transitions into early retirement (ATP, 2024). The early retirement scheme grants individuals the option to retire before the SRA, which they may use or forgo (receiving, in the latter case, a reimbursement of their contributions to the pension system); hence, payment into the early retirement scheme has an insurance value.

Occupational pensions are defined contribution (DC) and are primarily collective agreements based on bargaining between firms and labor unions or between firms and a pension company. This makes the occupational pensions quasi-mandatory, and their primary objective is to smooth income intertemporally by reallocating income from before to after retirement. Driven largely by the savings-based schemes, pension wealth in Denmark is almost 200% of GDP while the OECD average is 55%, see (OECD, 2024).

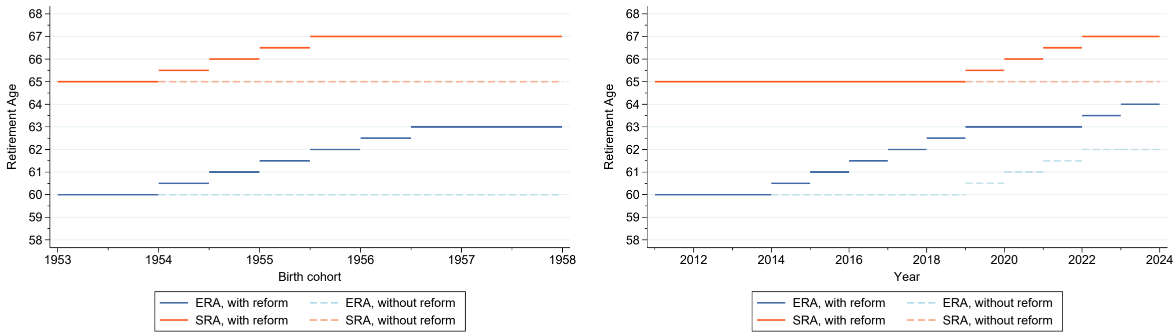
Private pensions are voluntary and offer the most individual flexibility for saving, also in relation to the payout profile. Private pensions also serve to offer retirement saving opportunities to those individuals not covered by the second pillar, such as the self-employed, and to people wanting to save more than the second pillar allows for.

2.2 The 2006 Welfare Agreement and the 2011 Retirement Reform

In 2006, the modern cornerstone of the sound public finances in Denmark, the "Welfare Agreement", was passed in parliament, aiming to counter the pressure on public finances of the expected increase in life expectancy by gradually increasing the ages for the early retirement age (ERA) and the statutory retirement age (SRA). The increases in the ERA and SRA were planned to start in 2019 and 2024, respectively. The reform that we use, the 2011 Retirement reform, was passed in parliament on May 13, 2011. It brought the increases 5 years forward, causing the first changes to take place in 2014. The main changes of the 2011 reform were to increase the ERA by ½ year every year from 2014 to 2017 and to increase the SRA by ½ year from 2019 to 2022. This increased the ERA from 60 years in 2014 to 62 years in 2017 and increased the SRA from 65 years in 2019 to 67 years in 2022. The 2011 reform also shortened the period between early and statutory retirement to a maximum of 3 years by further raising ERA, increasing it by half a year from 2018 to 2019 and again in 2022 and 2023. Figure 2 illustrates the changes in the early and statutory retirement ages. Figure 2a shows that all individuals born after 1954 had a biannual stepwise increase in both their ERA and SRA, while Figure 2b shows that these changes took effect from 2014 and 2019, respectively. The 2011 reform is also analyzed by García-Miralles and Leganza (2024b), who show that the reform led to increases in personal and employer-sponsored retirement contributions among the affected cohort, a mechanism argued to be due to "behavioral inertia", as individuals work longer, but continue to save.

Figure 2: The 2011 Reform's Effect on Retirement Ages

(a) Retirement Ages for Cohorts (b) Retirement Ages Over Time



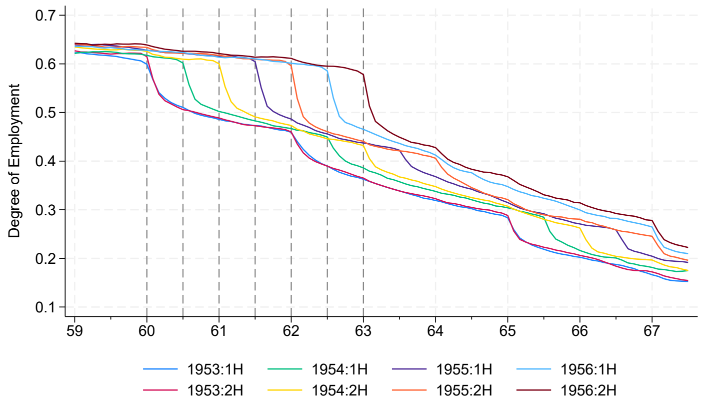
The figure illustrates the early retirement age (ERA) and statutory retirement age (SRA) for given cohorts (left) and in a given year (right). The bold lines illustrate the ERA and SRA under the rules introduced by the 2011 reform, whereas the dashed lines show what the ERA and SRA would be, absent the reform. The baseline corresponds to the ERA and SRA established by the 2006 Welfare Agreement.

Figure 3 illustrates that the 2011 reform was very effective in postponing the retirement decision for many workers. This can be seen as the employment degrees exhibit large dips of around 10 percentage points around the early retirement eligibility age, combined with the fact that these declines are postponed exactly as predicted by Figure 2 for the affected cohorts. The 10 percentage points dips in the employment degrees around the ERA also signify that only a relatively low share of workers go into early retirement. The curves in Figure 3 also exhibit dips two and five years after the ERA. These subsequent dips two years after the ERA for cohorts 1953-1955 arise because of a tax incentive yielding a tax-free bonus for individuals who postpone retirement until two years after their ERA.[10] For these

[10]For a full description of the tax-free bonus, see www.retsinformation.dk, § 74 m (in Danish). The bonus is a one-time payment corresponding to up to 6% of the annual early retirement benefits.

cohorts, the dips five years after the ERA are due to the statutory retirement age. The declines are also postponed as predicted by Figure 2.

Figure 3: Employment Degree for Cohorts Across Age



The figure shows the degree of employment for cohorts born in the years 1953-1956, born in the first (1H) or second half (2H) of the year. The last cohorts unaffected by the 2011 reform were born in 1953. The vertical, dashed, gray lines indicate the increased ERA for the cohorts (see Figure 2).

2.3 Data Sources

For our main sample, we use administrative register data for all individuals born in 1954 and 1955. To limit our sample to individuals affected by the 2011 reform covered in Section 2.2, we exclude individuals who (i) did not pay early retirement contributions at age 55, and (ii) are not of "Danish origin", as defined by Statistics Denmark.[11] Note that restriction (i) is imposed prior to the announcement of the 2011 reform to avoid endogeneity of the sample restrictions. Restriction (ii) is imposed due to a prior common practice of assigning a birthdate of either January 1 or July 1 to immigrants without information on their date of birth (Danish Ministry of Immigration and Integration, 2023), which would complicate the application of the RD design outlined in Section 3. Table B.1 illustrates how these two restrictions influence our main sample of individuals for a broad set of predetermined characteristics. Column (3) in Table B.1 describes the estimation sample for the main results in Section 4.

For our main health-related outcomes, we use two administrative registers that cover the healthcare utilization of all individuals in Denmark. These are (i) the Health Insurance Registry (SSSY), covering all contacts with GPs, dentists, physiotherapists, etc., at a weekly level, and (ii) the Prescription Drug Database (LMDB), which covers all prescribed drugs purchased at a daily level, with detailed information on the classification of medicine, following the Anatomical Therapeutic Chemical (ATC) classifi-

[11] Individuals of "Danish origin" are defined as having at least one parent with Danish citizenship born in Denmark. Individuals born in Denmark without available information about either parent are also considered of Danish origin.

cation system. We use the ATC codes to determine Defined Daily Doses (DDDs), cf. www.whocc.no. Using these two registers, we consider four main health-related variables: GP visits, antidepressant use, use of cardiovascular disease (CVD) medicine, and use of painkillers. For the latter three variables, we measure the extent of use by defined daily doses (DDDs). In Appendix D.1, we consider the extensive margin of usage/non-usage (binary) of these prescription drugs as well as the more severe health-related outcomes: Charlson Comorbidity Index (CCI), incidence of stroke, and mortality. To study CCI and strokes, we utilize the National Patient Register (LPR), covering all contacts with hospitals, the main reason for contact, and any diagnoses the patient has received upon contact. For mortality, we use whether an individual is present in the population register (BEF).

The choice of main health-related outcomes and their interpretation is described in detail in Section 2.3.1 below, and the definition of labor market outcomes is described in Section 2.3.2. We combine the health data with several other administrative registers. We obtain date of birth, gender, and civil status from the population register (BEF), income, net wealth, and payments to the early retirement age scheme from the income register (IND), education level from the education register (UDDA), labor market status from the Integrated Database for Labor Market Research (IDA), monthly employment data from The Employment Statistics of Employees (BFL), and public transfers on a weekly level from the Danish Research Institute for Economic Analysis and Modelling (DREAM) data.

2.3.1 Health and Healthcare Utilization Outcomes

This section describes the four health and healthcare utilization outcomes covered in Section 2.3: GP visits, antidepressants (DDDs), CVD medicine (DDDs), and painkillers (DDDs). It is important to note that the outcomes do not directly measure health, but are correlated with health. Objective health data is very hard to measure,[12] which is why we interpret the outcomes as a mix of health and healthcare utilization. For instance, the use of antidepressants is a function of factors such as mental health, detection, income, and preferences.

General Practitioner (GP) Visits: We consider GP visits as a general measure of healthcare utilization. As GP visits are free in Denmark, the primary cost incurred by individuals is the opportunity cost of time. While the other health-related outcomes are measured in half-yearly increments, GP visits are measured in yearly increments due to strong seasonality driven by closures during the summer and Christmas holidays and e.g., seasonal influenza (see also Figure B.5).

[12] Issues such as differences in detection rates across treatment and control groups are hard to rule out. While survey data can offer insights into self-assessed health, there is a body of literature that questions its reliability; e.g., Crossley and Kennedy (2002) find that 28% of respondents change their health status when asked twice in a survey.

Prescription Drug Usage: We analyze prescriptions for painkillers, antidepressants, and CVD medications to measure different dimensions of health and healthcare utilization that may be affected by an increased early retirement age. A key concern voiced in debates about increasing retirement ages is that physically demanding jobs can cause severe strain, sometimes requiring "worn-out" individuals to take pain medication to continue working.[13] Recently, attention has also turned to mental health, as jobs can potentially involve substantial psychological as well as physical demands.

The chosen prescriptions belong to two of the most common ATC code categories, C and N, with 46% and 33% of individuals aged 60 having a prescription for drugs for the cardiovascular system (ATC category C) and the nervous system (ATC category N) in 2010, respectively. See Appendix B.2 and www.whocc.no for more details on the ATC categories. 19% of individuals aged 60 had a prescription for painkillers (ATC N02, analgesics), and more than 11% had a prescription for antidepressants (ATC category N06A) in 2010. We specifically consider CVD medicine (ATC category C) to reflect that the risk of cardiovascular diseases is strongly connected to individuals' lifestyle factors that are related to (un)employment, including stress, and habits like diet, exercise, smoking, and drinking (Janlert et al., 1992, Eriksson et al., 2006, Caliendo et al., 2023). Furthermore, stress is known to lead to increases in allostatic load and telomere shortening, increasing the risk of cardiovascular diseases among other (McEwen, 1998, Seeman et al., 2001, Cawthon et al., 2003, Epel et al., 2004, Cutler et al., 2008). CVD medicine primarily consists of medicine to reduce and control blood pressure, such as renin-angiotensin system (RAS)-acting agents and beta and calcium channel blockers, and medicine that reduces cholesterol levels. See Figure B.4a for details. We consider the use of antidepressants as it has been shown to respond to retirement reforms (see, e.g., Grip et al. (2012), Carrino et al. (2020)). Similarly, we consider painkillers because previous literature has shown that the use of painkillers, especially opioids, is affected by economic conditions and unemployment (Pierce and Schott, 2020, Ahammer and Packham, 2023). Opioids and other analgesics and antipyretics are the most used drug types within painkillers, but antimigraine preparations are also included. See also Figure B.4c. While many non-opioid painkillers are sold as over-the-counter drugs, larger packages of painkillers (e.g., paracetamol and ibuprofen) have required a prescription since September 30, 2013.[14] Consequently, our measure of painkillers contains both strong and relatively weak treatments of pain. While consulting physicians to get prescriptions is free of charge and physicians have no financial incentive to prescribe certain brands or more expensive

[13]Denmark's largest union, 3F, whose members are primarily blue-collar workers, published a survey that 54% of its members had taken pain medication in the last year *due* to their job (3F, 2019).

[14]Sources (in Danish): www.laegemiddelstyrelsen.dk (1) and www.laegemiddelstyrelsen.dk (2). Packages of, e.g., more than 10 g of paracetamol or 4 g of ibuprofen require a prescription.

drugs, a caveat of our prescription outcomes is that they are still affected by income effects and individual price sensitivity. Leth-Petersen and Skipper (2014) show that demand for prescription drugs does vary with income for older individuals, and Simonsen et al. (2016) find some price sensitivity of demand for prescription drugs, but also that older individuals are less responsive to price changes.

2.3.2 Labor Market Outcomes

This section outlines the five labor market outcomes used to study the effect of the ERA increase on labor force participation. Following the International Labour Organization (ILO), employment takes precedence over unemployment, and unemployment takes precedence over being out of the labor force. We combine data from two sources: (i) monthly employment data from The Employment Statistics of Employees (BFL), and (ii) weekly data on public transfers from the Danish Research Institute for Economic Analysis and Modelling (DREAM) data.

Employment: We construct a binary indicator for employment, defined as having three or more months of full-time equivalent employment within a six-month period.

Other Transfers: We construct a binary indicator for receiving other transfers, defined as receiving public transfers (excluding retirement benefits) for at least three months within a six-month period while not being employed. We further disaggregate public transfers into *"Ordinary Transfers"* (unemployment insurance, cash benefits) and *"Health Transfers"* (sickness benefits, disability pension, etc.) (see Appendix B.3 for a detailed decomposition of other transfers).

Self-supporting: We construct a binary indicator for being self-supporting, defined as not being employed and not receiving any public transfers (including retirement benefits) concurrently for at least three months within a six-month period. As we have detailed data on transfers and earnings, this measure captures individuals who are relying on non-public pension income, savings, spousal support, or capital income for support.

3 Empirical Strategy

We exploit the discontinuous change in the early retirement age induced by the 2011 reform, outlined in Section 2.2, to identify the causal effects of increasing the retirement age on health and healthcare utilization outcomes. Specifically, we exploit that individuals born *on or after* July 1, 1954 (but before January 1, 1955) face an ERA of 61 years, whereas individuals born *before* July 1, 1954, (but after January 1, 1954) face an ERA of 60½ years. Likewise, individuals born after July 1, 1955, face an ERA of 62 years, and individuals born before July 1, 1955, face an ERA of 61½ years (see also Figure 2 for overview). We pool these two groups together to study the average effect of a half-year increase in the ERA.

We disregard the January-cutoffs also introduced by the reform. We choose July 1 as the focal point because of the large literature documenting the effects of school starting age on a vast number of outcomes (e.g., Dee and Sievertsen (2018) on mental health for Denmark and Arnold and Depew (2018) on long-run health in the US).[15]

We employ the local randomization framework formalized by Cattaneo and Titiunik (2022). As in Cattaneo et al. (2015), we assume there exists a window, $\mathcal{W} = [c-w, c+w]$, around the cutoff, c (July 1), where: (i) the distribution of the running variable, X_i, is the same for all individuals in $\mathcal{W}$: $F_{X_i|X_i\in\mathcal{W}(x)} = F(x)$, and (ii) potential outcomes for individuals in $\mathcal{W}$ depend only on the running variable, X_i, through treatment assignment: $Y_i(X_i) = Y_i(\mathbb{1}[X_i \geq c])$. Here, $\mathbb{1}[X_i \geq c]$ is an indicator variable equal to 1 if the individual is born on or after July 1. The existence of $\mathcal{W}$ implies that we can treat individuals near the cutoff as-if randomly assigned to the increased ERA. We choose the bandwidth around the cutoff, w, to be 50 days, meaning $\mathcal{W}$ is the interval from May 12 to August 20, being approximately 1.5 months on both sides of July 1. We choose this bandwidth using a data-driven approach, similar to Cattaneo et al. (2023). See Section 3.1 for details. Intuitively, it is plausible that individuals born on May 12 should not have significantly different health outcomes around retirement compared to individuals born on August 20, absent the 2011 reform.[16]

Estimating the discontinuous change in the ERA's effect on labor and health outcomes under the local randomization framework yields an RD design given by the following simple equation for individual i at relative time since ERA increase r:

$$Y_{ir} = \beta_{0,r} + \beta_{1,r} \cdot \mathbb{1}[X_i(t) \geq c(t)] + Z_i\delta_r + \varepsilon_{ir}, \tag{1}$$

where Y_{ir} is a given labor or health-related outcome for individual i at relative time since ERA increase r. The cutoff, $c(t)$, depends on the specific cohort t, and the running variable, $X_i(t)$ depends on the distance to the cohort-specific cutoff. The cohort-specific cutoffs, $c(t)$, are given as July 1, 1954, for the 1954:2Q-1954:3Q cohort and July 1, 1955, for the 1955:2Q-1955:3Q cohort. $\beta_{0,r}$ is a constant term, and $\mathbb{1}[X_i \geq c]$ is an indicator variable equal to 1 if the individual is born on July 1 or after with the associated coefficient, $\beta_{1,r}$, being of main interest. If the effect of increasing the ERA by ½ year is

[15]In Appendix D.2, we show that the majority of estimated effects are similar, independent of the cutoff used.

[16]In Appendix C.4, we show that our main results are robust to different bandwidth choices.

homogeneous across the two cohorts, the pooling strategy will estimate the same effect as estimating them separately, but with higher precision due to the larger sample. In the presence of heterogeneity, e.g., health effects increasing in age, we instead estimate the average treatment effect across cohorts, which may be higher or lower than the cutoff-specific effects. Z_i is a vector of control variables[17] included to increase precision. ε_{ir} is the idiosyncratic error term. We use conventional robust standard errors for inference.[18]

For our main specification, we measure our health-related outcomes for a two-year period following the ERA increase (age 60½-62½ for the 1954:2Q-1954:3Q cohort and 61½-63½ for the 1955:2Q-1955:3Q cohort). Consequently, we compare health outcomes of individuals at a given *relative* time since the ERA increase, but at different points in time. E.g., for someone born July 11, 1954, we measure painkiller DDDs from January 11, 2015 (at age 60½) to January 10, 2017 (the day before turning 62½). By measuring the outcomes at specific ages rather than at specific dates, we implicitly allow for age effects on health outcomes and assume time effects are zero in $\mathcal{W}$. The assumption that time effects are zero follows from the local randomization framework, i.e., that potential outcomes are not affected by the running variable for values in $\mathcal{W}$. Note that the local randomization framework does formally impose the non-existence of a time gradient in health. However, while this is most likely not the case literally, we argue that any potential time gradient is approximately zero in $\mathcal{W}$, so only a very small part of our estimate reflect this. In Appendix C.3, we show that using the continuity framework, where the running variable is allowed to affect the outcome directly, would lead to similar estimates, albeit with much larger confidence intervals, as expected.

Note, $\beta_{1,r}$ in Equation (1) can be interpreted as the intent-to-treat effect (ITT) of the reform. If one expects that the reform will only affect individuals' health and healthcare utilization through increased employment, one could estimate an IV equation instead, by instrumenting employment status with the dummy for being born after July 1. However, this would require an exclusion restriction, i.e., that the reform affects health *only* through employment. The exclusion restriction would be violated if, for instance, the affected individuals feel unfairly treated by the reform, increasing the risk of depression independent of extending employment. For the main part of our analysis, we essentially estimate the reduced form equation and thereby avoid the exclusion restriction.

[17]Control variables are measured at age 55 if not otherwise specified and include: a male dummy, a marriage dummy, dummy for short cycle tertiary education or higher, average net wealth (excluding pension wealth) age 48 to 52 squared, average disposable income age 48 to 52 squared, and a dummy for being born in 1955 (1954 as the reference group). All variables are measured prior to the reform announcement to avoid endogeneity.

[18]With over 23,000 individuals in our main specification, we comfortably rely on asymptotic inference. An alternative would be to use Fisherian inference as suggested by Cattaneo et al. (2023), which would be valid in any finite sample.

Using the local randomization framework over the more common continuity-based framework significantly increases statistical power, as it does not require approximating a potentially complex functional form of the potential outcome in the limit at the cutoff. Instead, we assume a constant relationship between the running variable and potential outcomes on both sides of c. This also allows us to use all values of the running variable in $\mathcal{W}$ equally, i.e., without using weights. Consequently, standard errors on our estimates are smaller, and we can reject effects that are much larger or smaller than our main estimate. This is important, since potentially large adverse health and healthcare utilization effects would have important policy implications, even if not statistically significant.

3.1 Window Choice

We choose the bandwidth around the cutoff, $w = 50$, using a data-driven approach similar to Cattaneo et al. (2015, 2023). We choose a set of predetermined health and healthcare utilization covariates and test the presence of a treatment effect at the cutoff for these predetermined covariates. Intuitively, individuals born just around the cutoff should be completely comparable (e.g., individuals born on July 1 and June 30), and increasing the bandwidth leads to increasingly differing individuals because of, e.g., different health status by season of birth. Formally, we test the difference in means of the following health-related variables: GP visits, usage of painkillers, antidepressants, and CVD medicine (measured both by binary indicators of use and by defined daily doses), incidence of stroke, and the Charlson Comorbidity Index (CCI). We measure all predetermined health and healthcare utilization variables for a four-year period prior to the announcement of the reform (age 52½-56½ for 1954:2Q-1954:3Q cohort and age 51½-55½ for 1955:2Q-1955:3Q cohort). By choosing these age intervals, the predetermined variables are completely exogenous to the reform.

The resulting P-values are depicted in Figure C.1a. Cattaneo et al. (2023) suggest using the largest possible window with the minimum P-value above 0.15, this level indicated by the dashed horizontal line in Figure C.1a. Figure C.1b illustrates the absolute value of the standardized difference as a function of the window choice. An absolute standardized difference above 0.25, indicated by the horizontal line in Figure C.1b, can be interpreted as a sign of imbalance (Stuart and Rubin, 2008). Figure C.1 clearly shows that most P-values from testing the predetermined covariates are relatively high, and the standardized differences in predetermined covariates are very low for all bandwidth choices. We choose a bandwidth of $w = 50$, because the P-values testing the difference in antidepressants DDDs start declining at higher bandwidths. We further test the sensitivity of our results to this choice in Appendix C.4 where we show that our results are overall quite robust to increasing or decreasing the bandwidth.

4 Results

In this section, we first document the 2011 reform's effect on employment, transitions into public transfers other than early retirement, and the probability of being self-supporting in Section 4.1. Intuitively, for the reform to affect health and healthcare utilization, a necessary condition is that it affects labor force participation, otherwise the increased ERA is not a binding constraint for the targeted population.

Second, we examine any potential health and healthcare utilization consequences of the increased ERA in Section 4.2. In Section 4.2.1, we examine the robustness of our results using a large set of alternative specifications. Lastly, in Section 4.3, we discuss additional health and healthcare utilization outcomes, and in Section 4.4, we consider cutoff-specific effects.

4.1 The Increased ERA's Effect on Labor Force Participation

In Table 1, we report estimates from the regression specified in Equation (1), both with and without control variables. While including control variables is not necessary for identification in our local randomization design, we include controls to increase precision and to confirm the stability of our results. The outcome variables are different measures of labor force participation measured in the half-year following the ERA extension, i.e., the period during which individuals born on or after July 1 were ineligible for early retirement, while those born before were eligible. The corresponding estimates without controls are illustrated in Figure 4, which presents a graphical version of the regression using binned outcomes. In the following, we focus on Panel B in Table 1, including controls. Column (1) confirms the points illustrated by Figure 3: Increasing the ERA by ½ year leads to an increase in employment of 17 percentage points in the half-year period following the ERA increase, corresponding to 27%.

However, the reform also led to an increase in individuals transitioning into other transfers of 4 percentage points (41%), as seen from Column (2). In Columns (3) and (4), we decompose "other transfers" into ordinary transfers and health transfers.[19] Ordinary transfers almost exclusively consist of unemployment insurance and cash benefits, while health transfers consist largely of disability pension and sickness benefits. A large fraction of the increase in the transition into other transfers (Column (2)) stems from an increase in ordinary transfers (Column (3)). Contrarily, the effect on health transfers (Column (4)) is not statistically significant at the 10% level. This is for outsiders perhaps surprising, but not unexpected given that the Danish flexicurity system makes ordinary transfers easily accessible, whereas health transfers typically require a formal health assessment, making them less readily available.

[19] See Appendix B.3 for further decomposition of "other transfers".

We also show an increase in the number of self-supporting (Column (5)), i.e., individuals who are neither employed nor receiving any government transfers of 2 percentage points (26%). Given that the ERA increase was only a half-year increase, individuals could choose to use their savings to bridge the period until early retirement eligibility. In Appendix B.4, we plot the average share of individuals who are employed or receive other transfers over time separately for those born before or after the cutoff date. The increase in transfers is partly mechanical, as some individuals on transfers who would have transitioned to early retirement instead remain on other transfers due to the increased ERA. In Appendix C.2 we condition on employment in the half-year period prior to the ERA increase to limit mechanical effects on other transfers and self-support. We show that conditional on prior employment, there is still an increase in self-supporting, but not in other transfers. Despite this, there is still an increase in *the ratio of time spent* on mainly ordinary transfers, as well as health transfers. The effect on exiting employment can be driven by employees quitting their jobs and/or employers firing the employees at the extended ERA. Regrettably, the available data do not allow us to distinguish between the effect caused by voluntary resignations and that caused by involuntary terminations.

Table 1: Effects on Labor Force Participation

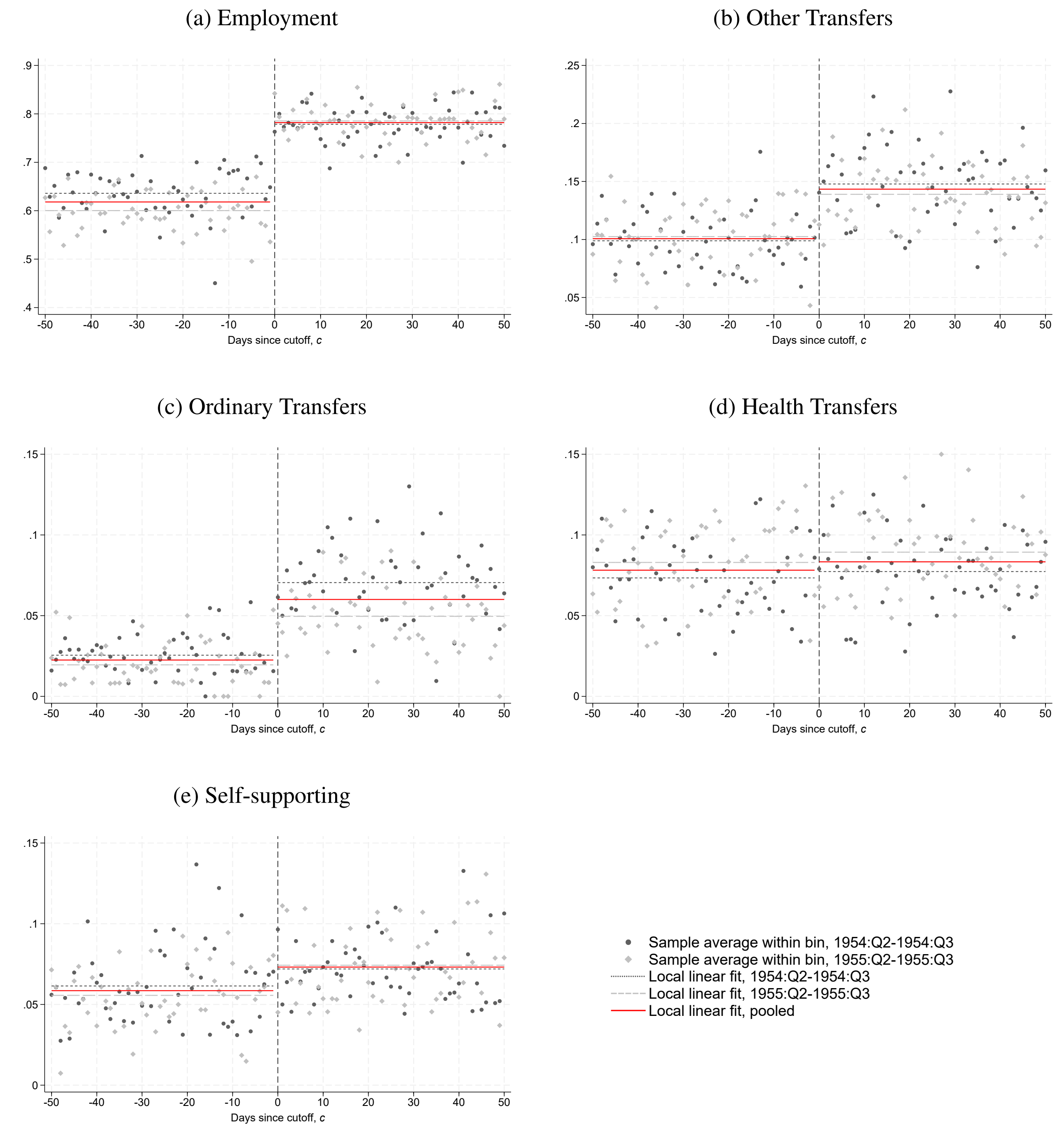
	(1) Employment (0/1)	(2) Transfers (0/1)	(3) Ordinary transfers (0/1)	(4) Health transfers (0/1)	(5) Self-supporting (0/1)
Panel A: No controls					
$\mathbb{1}[X_i \geq c]$	0.163***	0.0423***	0.0375***	0.00475	0.0151***
	(0.00585)	(0.00426)	(0.00259)	(0.00354)	(0.00323)
Pct. Change	26.43	41.78	165.25	6.05	25.78
Panel B: With controls					
$\mathbb{1}[X_i \geq c]$	0.165***	0.0412***	0.0373***	0.00394	0.0154***
	(0.00575)	(0.00421)	(0.00258)	(0.00351)	(0.00320)
Pct. Change	26.70	40.72	164.11	5.03	26.41
Mean	0.62	0.10	0.02	0.08	0.06
N	23,694	23,694	23,694	23,694	23,694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on labor force participation outcomes estimated using Equation (1). All outcomes are measured in the first half-year following the ERA increase (age 60½-61 for 1954:2Q-1954:3Q cohort and age 61½-62 for 1955:2Q-1955:3Q cohort). See Section 2.3.2 for a detailed description of labor market outcomes. "Mean" refers to the average for the group born before July 1 (i.e., prior to the cutoff). Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

Danish Economic Councils (2021) find a similar increase in employment caused by the 2011 reform, using a comparable RD strategy. Athey et al. (2020) also examines the impact of the 2011 reform on employment and benefit receipt, while García-Miralles and Leganza (2024b) studies its effects on early

Figure 4: Effects on Labor Force Participation

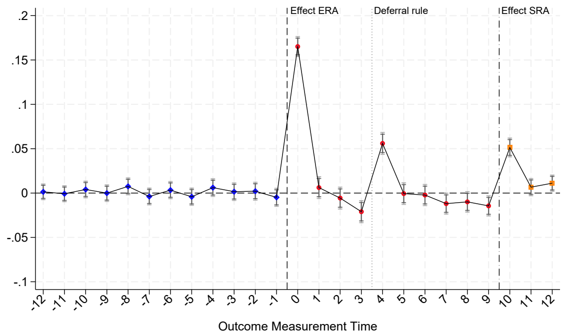


The figure illustrates the results from Table 1, estimating Equation (1) on labor force participation outcomes. Figures 4a to 4e correspond to Columns (1) to (5) in Table 1. The figures show averages of the outcome variables for each day of the running variable for 50 days on each side of the cutoff by birth cohort. "Local linear fit" refers to the average within the bandwidth of 50 days on each side of the cutoff date. All outcomes are measured in the first half-year following the ERA increase (age 60½-61 for 1954:2Q-1954:3Q cohort and age 61½-62 for 1955:2Q-1955:3Q cohort).

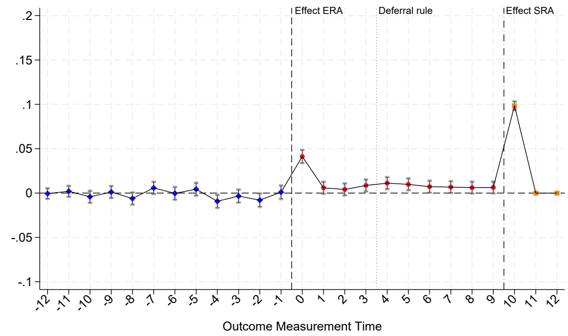
retirement benefits claiming and retirement, both finding similar patterns. The increased ERAs' effect on employment in Column (1) of Table 1 is also similar to the findings of, e.g., Mastrobuoni (2009), Manoli and Weber (2016b), while the effects on substitution into other transfers in Columns (2)-(4) are similar to

Figure 5: Dynamic Effects on Labor Force Participation

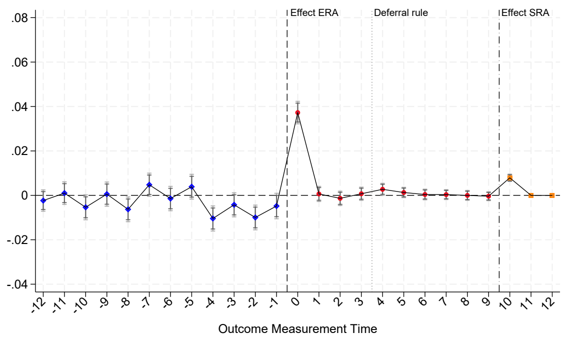
(a) Employment



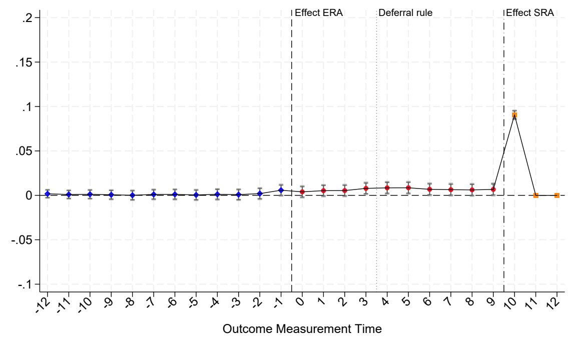
(b) Other Transfers



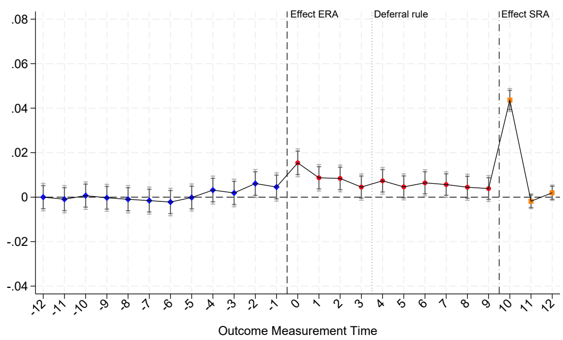
(c) Ordinary Transfers

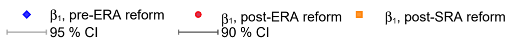


(d) Health Transfers



(e) Self-supporting





The figure illustrates the results from estimating Equation (1) on labor force participation outcomes in half-year intervals from 6 years before the ERA increase to 6.5 years after. See Section 2.3.2 for a detailed description of labor market outcomes. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and the average of net wealth and disposable income (ages 48–52), including squared terms. The gray capped spikes show 90% and 95% confidence intervals calculated with robust standard errors.

the findings of e.g. Duggan et al. (2007), Vestad (2013). In addition, the results summarized by Table 1 very closely resemble those of Staubli and Zweimüller (2013), who find that raising the ERA in Austria increased employment by 9.75 and 11 percentage points for the affected men and women, respectively, and sizable spillover effects on employment insurance programs, but negligible effects on disability insurance claims. Table 1 and Figure 4 make the 2011 reform's effect on labor force participation very apparent, but also highlight that some individuals bridge the gap to retirement by using transfers other than early retirement or by self-financing early retirement. That some individuals bridged the gap to retirement by taking up other transfers is also found by Athey et al. (2020).

We also show the dynamic effects of the ERA increase on labor force participation in Figure 5 using half-year intervals where $t = 0$ denotes the period immediately following the ERA change. The analysis spans from $t = -12$ (6 years prior to the ERA increase) to $t = 12$ (6.5 years after the ERA increase). For employment, we observe distinct jumps at $t = 0$ (corresponding to the ERA increase), $t = 4$ (deferral rule), and $t = 10$ (SRA increase), consistent with the patterns in Figure 3. We also observe an increase at $t = 10$ for other transfers (ordinary and health transfers) and the probability of being self-supporting due to the half-year increase in the statutory retirement age.

4.2 The Increased ERA's Effect on Health and Healthcare Utilization

In Table 2, we estimate Equation (1) to identify the 2011 reform's effect on health and healthcare utilization. In our main specification, we measure health-related outcomes over a two-year period after the extended ERA (i.e., age 60½-62½ for the 1954:2Q-1954:3Q cohort and age 61½-63½ for the 1955:2Q-1955:3Q cohort). Figure 6 shows the graphical equivalent without controls. While we examine the half-year period directly affected by the ERA increase for labor supply, the timing of potential health effects is less certain, as they may not emerge immediately. Therefore, our main specification uses a two-year window to allow health effects to materialize over time. We also consider half-year, four-year, and six-year periods in Table 3. In addition, we present dynamic health outcomes by half-year intervals (by year for GP visits due to seasonality), covering a range from 8 years before to 6.5 years after the ERA increase in Figure 7.[20]

The estimates in Column (1) of Table 2 and Figure 6a show no effect on GP visits, and we can reject increases of more than 2% at the 5% level. Column (2) of Table 2 and Figure 6b also indicate no effect on painkillers, measured by DDDs, and again we can reject large increases of more than 11% at the 5% level. However, we find a borderline statistically significant positive effect on antidepressants in Column (3) and a borderline insignificant positive effect on CVD medicine in Column (4). For antidepressants, we find an increase of 6 DDDs (corresponding to an increase of 9%). This effect is small

[20] As the 1954:2Q-1954:3Q and 1955:2Q-1955:3Q cohorts experience the announcement of the reform at different points relative in time to their ERA increase, we do not distinguish between pre- and post-announcement estimates in these graphs, in contrast to the cutoff-specific dynamic effects in Appendix D.2.

Table 2: Effects on Health and Healthcare Utilization Outcomes

	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
Panel A: No controls				
$\mathbb{1}[X_i \geq c]$	-0.0111	2.009	5.898*	30.18*
	(0.0943)	(2.971)	(3.487)	(16.73)
Pct. Change	-0.15	2.98	9.35	3.99
Panel B: With controls				
$\mathbb{1}[X_i \geq c]$	-0.0336	1.547	5.803*	23.95
	(0.0936)	(2.967)	(3.475)	(16.50)
Pct. Change	-0.46	2.29	9.20	3.17
Mean	7.36	67.46	63.09	756.07
N	23,694	23,694	23,694	23,694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

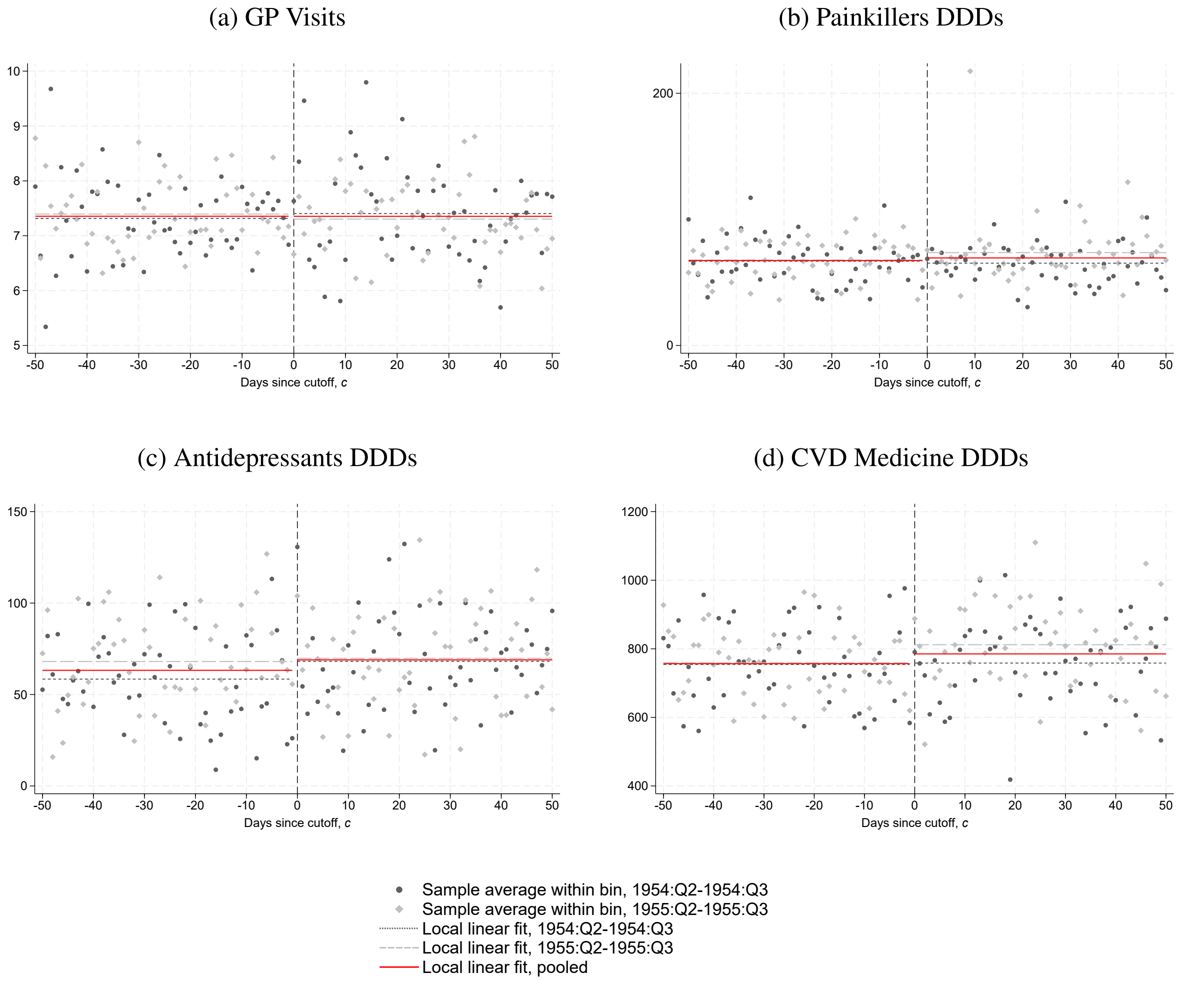
Effects on health and healthcare utilization outcomes estimated using Equation (1). All outcomes are measured over a two-year period after the ERA increase (age 60½-62½ for the 1954Q:Q2-1954:Q3 cohort and age 61½-63½ for the 1955Q:Q2-1955:Q3). The outcome variables are described in detail in Section 2.3.1 and the frequencies of prescriptions more generally are described in Figures B.3 and B.4. "Mean" refers to the average for the group born before July 1 (i.e., prior to the cutoff). Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

relative to prior evidence. For example, Grip et al. (2012) show that a Dutch pension reform increased depression rates by 40% for men, and Serrano-Alarcón et al. (2023) find that an Italian pension reform increased mental health hospitalizations by 69% for women. Our estimate is closer to, but still lower, than Carrino et al. (2020) that finds an increase of 19% in mental health disorders for women in UK when increasing the state pension age. In contrast, Hagen (2018) finds no effect on increasing the retirement age on mental drugs for female local government workers.[21] Overall, the 95% confidence intervals for percentage changes reported in Panel B range from (–2.95%, 2.04%) for GP visits, (-6.33%, 10.91%) for painkillers, (–1.60%, 19.99%) for antidepressants, and (–1.11%, 7.45%) for CVD medicine. Figure 7c illustrates that the effect on antidepressant DDDs builds up from the half-year before the extended ERA, but shrinks drastically after around 4½ years.[22] In the specification that does not include covariates, we find a borderline statistically significant effect on CVD medicine of 30 DDDs (4% increase). However, including covariates, the estimate becomes smaller (24 DDDs) and insignificant at the 10% level. Yet,

[21] It should be noted that the estimates to which we compare our results are based on gender-specific pension reforms, whereas our estimates capture an average effect for both men *and* women.

[22] However, we caution that at older ages, we measure outcomes in years that are also partly affected by the Covid-19 pandemic.

Figure 6: Effects on Health and Healthcare Utilization Outcomes



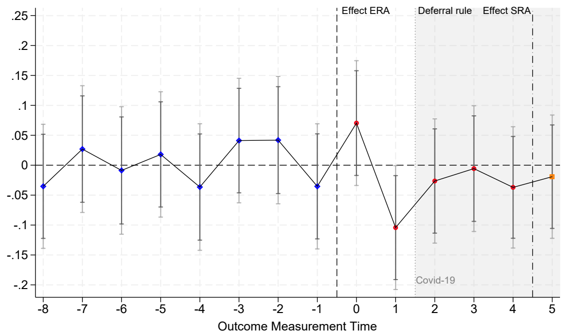
The figure illustrates the results from Table 2, estimating Equation (1) on health and healthcare utilization outcomes. Figures 6a to 6d correspond to Columns (1) to (4) in Table 2. The figures show averages of the outcome variables for each day of the running variable for 50 days on each side of the cutoff. "Local linear fit" refers to the constants fitted on the data within the bandwidth of 50 days on each side of the cutoff date, July 1. All outcomes are measured over a two-year period following the ERA increase.

in Figure 7d, there is some indication of a positive effect on CVD medicine from $t = -1$ to $t = 0$, with a statistically significant effect observed at $t = 4$.

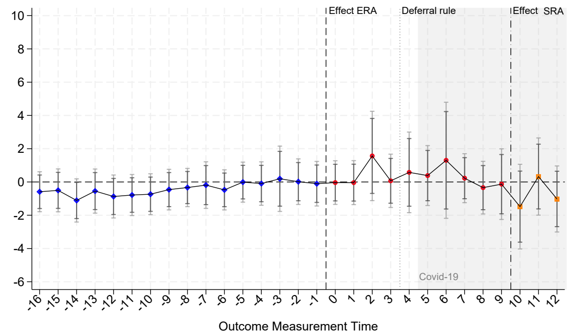
In Table 3, we present the effects of the ERA increase across several periods: four years prior to the announcement (Panel A), the entire post-announcement period (Panel B), and the periods following the ERA increase from short-run effects (half-year intervals for prescription drugs, yearly for GP visits in Panel C), as well as longer-term effects over four years (Panel D) and six years (Panel E). These effects are basically aggregations of the effects shown in Figure 7. We find no effects in the pre-announcement period, consistent with the expectation that individuals born around the cutoff date should be fully comparable prior to the announcement of the reform. We, however, caution that for antidepressants, while

Figure 7: Effects on Health and Healthcare Utilization Outcomes Over Time

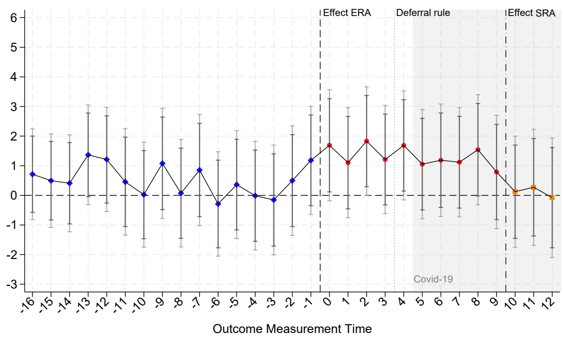
(a) GP Visits



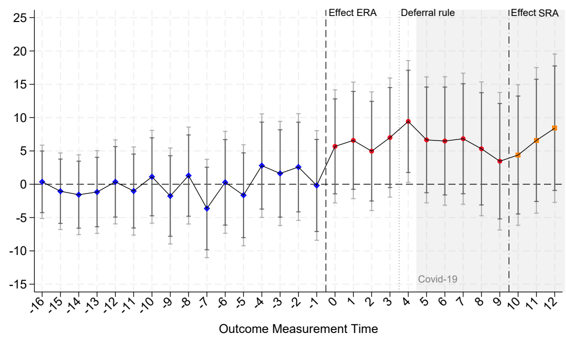
(b) Painkillers DDDs

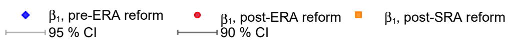


(c) Antidep. DDDs



(d) CVD Med. DDDs





The figure illustrates the results from estimating Equation (1) on main health and healthcare utilization outcomes in half-year intervals (yearly intervals for GP visits) from 8 years before the ERA increase to 6.5 years after. Outcome variables are described in detail in Section 2.3.1. The gray bar indicates that the estimates include years affected by the Covid-19 pandemic, which may influence healthcare utilization. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms. The gray capped spikes show 90% and 95% confidence intervals calculated with robust standard errors.

not statistically significant, the difference in the pre-announcement period is 7 DDDs (7%). We also find no statistically significant post-announcement effects. Post-ERA increase, we find no effects on GP visits and painkillers for any time period. For antidepressants, we find a borderline statistically significant effect in the short run (1.7 DDDs, an 11% increase). This is the period of time during which our treatment group is not eligible for early retirement, but the control group can retire early. We find no significant effects for a longer time period. For all post-ERA periods for CVD medicine, we find an increase of around 3%, though not statistically significant.

Overall, our results suggest that the increase in the ERA had no measurable impact on GP visits

Table 3: Effects on Health and Healthcare Utilization Outcomes

	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
Panel A: Pre-announcement (4 years)				
$\mathbb{1}[X_i \geq c]$	0.0793	-5.249	7.326	-6.967
	(0.165)	(4.944)	(5.668)	(21.79)
Mean	13.93	48.96	105.85	777.61
Pct. Change	0.57	-10.72	6.92	-0.90
N	24,286	24,286	24,286	24,286
Panel B: Post-announcement				
$\mathbb{1}[X_i \geq c]$	-0.108	-3.058	1.356	9.123
	(0.207)	(5.151)	(8.165)	(34.64)
Mean	17.74	100.40	165.80	1484.68
Pct. Change	-0.61	-3.05	0.82	0.61
N	24,237	24,237	24,237	24,237
Panel C: Post-ERA (short-run)				
$\mathbb{1}[X_i \geq c]$	0.0704	-0.0377	1.691*	5.684
	(0.0532)	(0.669)	(0.957)	(4.329)
Mean	3.62	16.39	16.05	178.84
Pct. Change	1.95	-0.23	10.53	3.18
N	23,694	23,694	23,694	23,694
Panel D: Post-ERA (4 years)				
$\mathbb{1}[X_i \geq c]$	-0.0639	4.057	10.81	52.84
	(0.166)	(6.811)	(6.694)	(33.23)
Mean	14.52	143.89	125.46	1636.62
Pct. Change	-0.44	2.82	8.61	3.23
N	23,694	23,694	23,694	23,694
Panel E: Post-ERA (6 years)				
$\mathbb{1}[X_i \geq c]$	-0.112	2.529	13.47	71.88
	(0.229)	(9.196)	(9.785)	(50.56)
Mean	21.12	232.79	191.15	2682.38
Pct. Change	-0.53	1.09	7.05	2.68
N	23,694	23,694	23,694	23,694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on health and healthcare utilization outcomes estimated using Equation (1). In Panel A, outcomes are measured for a four-year period prior to the announcement of the reform, and in Panel B, outcomes are measured post-announcement and pre-ERA increase (age 56½-60½ (four years) for 1954:Q2-1954:Q3 cohort and age 55½-61½ (six years) for 1955:Q2-1955:Q3 cohort). In Panel C, outcomes are measured for a half-year period after ERA increase (one year for GP visits), in Panel D, outcomes are measured for a four-year period after ERA increase, and in Panel E, outcomes are measured for a six-year period after ERA increase. The outcome variables are described in detail in Section 2.3.1 and the frequencies of prescriptions more generally are described in Figures B.3 and B.4. "Mean" refers to the average for the group born before July 1 (i.e., prior to the cutoff). Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

or painkiller use. This is in contrast to Nielsen (2019), who finds a reduction in GP visits after early retirement for Denmark. Nielsen compares individuals younger than 60 years old with those aged 60 or older who were eligible for early retirement benefits (ERA of 60 years in Denmark prior to the 2011 reform) using an RD design. Nielsen argues that if employers grant employees time off work to visit the doctor, there is no leisure cost of going to the GP's office before retirement. However, Nielsen is comparing individuals across different age groups, and thus, it is hard to separate age health effects from the effects of ERA. Similar to us, Cremers et al. (2024) examine the 2011 reform, but focus on the effect of early retirement itself, rather than delaying retirement, on the probability of prescription drug use. They find that early retirement decreases the probability of using painkillers but increases the likelihood of hypertension and mental health medication use. However, Cremers et al. (2024) use a window of 4 years around January 1, 1954 (January 1, 1950 to December 31, 1957), whereas we use a window of 50 days. In a sensitivity analysis, when Cremers et al. (2024) reduce their window to 1 year, they still find a reduction in painkillers, but no statistically significant effects of retirement on hypertension or mental health.

4.2.1 Alternative Specifications and Robustness Checks

Using the Continuity Framework: The local randomization framework imposes that potential outcomes for individuals in the window $\mathcal{W}$ only depend on the running variable through treatment assignment. In our setting, as discussed in Section 3, this assumption implies that any health gradient in time is negligible in $\mathcal{W}$. In Appendix C.3, we loosen this assumption by employing the continuity framework instead, which allows the running variable to affect potential outcomes. We use the mean squared error (MSE) optimal bandwidth with triangular weights, as proposed by Calonico et al. (2014) and Cattaneo et al. (2019) and report both conventional and robust biased-corrected estimates. The results reported in Table C.2 are quantitatively very similar to those in Table 2, but with much larger standard errors. The one exception is for CVD medicine, where we now observe a negative estimate, however, the robust bias-corrected confidence interval spans from -155 to 96 DDDs, placing our estimate in Table 2 well within its bounds. The larger standard errors are a result of the smaller bandwidths, using triangular weights, and the additional information needed to estimate the local polynomials. Given that estimates are overall quantitatively similar to our main specification, we favor the local randomization framework because of the large increase in precision it offers.

Choice of Bandwidth: In Section 3.1, we chose the bandwidth $w = 50$ using a data-driven approach. In Appendix C.4, we examine the sensitivity of our results to this choice by increasing and decreasing the bandwidth. Overall, our main results are very robust to the choice of bandwidth, as shown in Figure C.3.

Sensitivity to Observations Close to the Cutoff: Bajari et al. (2011) and Barreca et al. (2011,

2016) highlight that heaping around the cutoff may introduce bias in RD designs. Heaping is unlikely to affect our results as we use date of birth as the running variable, which is hard, if not impossible, to manipulate (see Figure B.1), and we do not assign higher weights to observations close to the cutoff. In Appendix C.5, we show that our results are indeed very robust to excluding observations near the cutoff by employing "donut-hole" regressions of varying radiuses.

Sensitivity to Outliers: We also test the sensitivity of our results to outliers by winsorizing our main health and healthcare utilization outcomes at the 1^{st} and 99^{th} percentile. We show the results in Appendix C.6, and we find that our results remain similar to Table 2.

Regression Discontinuity Difference-in-Differences (RD-DD): In Section 3, we argued that using January 1 as a cutoff could have implications for our results, as this cutoff also affected school starting age. If school starting age (or other discontinuities that may be present at the cutoff) has a positive effect on health, our empirical strategy in Section 4.2 may capture both the health and healthcare utilization effects of the increased retirement age and those related to differences in school starting age. Similarly, the presence of any other discontinuities around July 1 will bias our results in Section 4.2. To explore the presence of such biases, we employ an RD-DD strategy in Appendix C.7. We use the cohorts born in 1952:2Q-1952:3Q and 1953:2Q-1953:3Q (unaffected by the reform) as control cohorts to wash out any general post-July 1 effect of our estimate of interest. Table C.4 and Figure C.5 are very similar compared to Table 2 and Figure 6, which makes the presence of a general July 1 effect unlikely. The main difference is that the effect on GP visits changes sign, but the estimate is still close to zero. Additionally, we employ an RD-DD design using individuals who did not pay early retirement contributions at age 55 as our control group. This group is not eligible for early retirement and is hence unaffected by the 2011 reform. Since individuals who did or did not contribute to the early retirement scheme can be observed at the same age and point in time concurrently, this approach enables us to account for time-specific shocks, such as flu pandemics. Our results in Table C.5 and Figure C.6 show the results. First, we demonstrate that individuals who did not contribute to the early retirement scheme at age 55 exhibit higher healthcare utilization on both sides of the cutoff compared to those who did contribute. Second, while the estimates are not statistically significant due to larger standard errors, we find for CVD medicine an estimate of 40 DDDs (a 5% increase), while for antidepressants, the estimate is -1 DDD, though the confidence interval ranges from -17 to 16 DDDs (a range including our main estimate of 6 DDDs). We find the result for the antidepressants particularly informative, since the estimate was borderline significant in the main specification, but with a relatively large (but insignificant) pre-announcement effect. Thus, this could indicate that some of the effect on antidepressants in the main specification is driven by pre-announcement differences, but these are taken into account when we use a control group, and thus, the estimate of antidepressants should fall, which it indeed does.

Overall, our results are somewhat sensitive to the specifications in Section 4.2.1. For painkillers and GP visits, we consistently find small and insignificant effects, though with varying precision. In contrast, the evidence for antidepressants and CVD medicine is less robust. While the main specification suggests borderline significant positive effects, alternative specifications yield negative estimates (–1 DDD for antidepressants in the RD-DD with non-ERA contributors, and –30 DDDs for CVD medicine under the continuity framework). Taken together, we find small or no effects on health and healthcare utilization of the reform.

4.3 Additional Health and Healthcare Utilization Outcomes

While our primary health-related outcomes are GP visits and prescription drug usage measured by DDDs, we also consider a binary indicator for use of prescription drugs in Appendix D.1 as well as the more severe and objective health outcomes: Charlson Comorbidity Index (CCI), incidence of stroke, and mortality (see also Appendix D.1 for description). We find no effect on the binary indicators for painkillers and antidepressants usage for any time period, but we do find a positive effect on the probability of CVD medicine usage in the post-announcement period. We find that a half-year increase in the ERA raises the probability of taking CVD medication by 1.3 percentage points (a 3% increase) following the announcement of the reform, but before the half-year period without early retirement eligibility begins.

We find no effects of the half-year increase in the ERA on CCI or mortality. For stroke, we observe a borderline statistically significant reduction in the probability of experiencing a stroke by 0.001 percentage points in the half-year following the ERA change. However, given that stroke is a rare health outcome, especially over a short period of time, this estimate is sensitive and should be interpreted with caution. When extending the analysis to a two-year period, we instead find a small, statistically insignificant increase of 0.001 percentage points on the probability of having a stroke. Given the limited effects on less severe health-related outcomes such as GP visits and prescription drug use, the absence of significant impacts on more severe health outcomes, such as CCI and mortality, is consistent with our expectations.

Overall, these results suggest that the effect of the half-year increase in the ERA on additional health and healthcare outcomes is limited, with no consistent impact across most outcomes except for a small, short-term increase in the probability of having a CVD medicine prescription following the reform announcement.

4.4 Cutoff-Specific Effects

One reason why we do not find large effects could be that the individuals affected by the reform are too young. To test this, we estimate the cutoff-specific effects in Appendix D.2. We start with the two summer-cutoffs included in our pooled main specification: 1954:Q2–1954:Q3 and 1955:Q2–1955:Q3. For these cohorts, we present both a table of effects across different periods (Table D.3 and Table D.4) and dynamic effects over time (Figure D.2 and Figure D.3). We find that the effect on antidepressant use is primarily driven by the 1954 cohort, whereas the effect on CVD medicine use is driven by the 1955 cohort. The differences in effects between the two cohorts make a consistent causal effect of an increasing ERA less convincing and could indicate that the results may be due to cohort-specific factors or random variation.

Second, we estimate the effect of increasing the ERA by ½ year for the winter-cutoffs in two separate periods: Table D.5 (1953:Q4-1954:Q1) and Table D.6 (1954:Q4-1955:Q1). While these are not included in our main specification due to the large literature documenting school starting effects, we do not find any effects on health-related outcomes for any periods (pre-announcement, post-announcement, post-ERA), except for a negative effect on GP visits in the short-run for the 1954:Q4-1955:Q1-cohort. Hence, these results are consistent with there being no overall effect of increasing the ERA by ½ year on health-related outcomes for the overall population.

Last, it is also possible that the effect of increasing the ERA does not materialize at younger ages (e.g., from 60 to 60.5 years as for the 1953:Q4-1954:Q1 cohort) but may emerge at later ages. As a final check, we therefore examine the 1956:Q2–1956:Q3 cohort to assess the impact of raising the ERA from 62.5 to 63 years in Table D.7. However, it is important to note that for this group, the changes in the deferral rule and the statutory retirement age (SRA) differ compared to the other groups[23] – one reason this cohort is not included in the pooled specification. Additionally, our ability to study long-term health-related outcomes for this group is more limited, as they are younger, and the ERA increase occurs at a later age. For this group, we find no effects on any main health-related outcomes in the pre- and post-announcement period. However, for both the short-run and two-year periods after the ERA increase, we observe a statistically significant increase in CVD medicine usage. In the short run, i.e., the half-year period following the extended ERA, this corresponds to an increase of 16 DDDs (an 8%

[23] For the 1956 cohort, the period between the ERA and the SRA is reduced to 4.5 years, compared with 5 years for earlier cohorts, and the period between the ERA and deferral rule is shortened from 2 years to 1.5 years.

increase). We find no statistically significant effects on antidepressants for this group.

5 Mechanisms

So far, the main results indicate a small, if any at all, effect of the 2011 reforms on health and healthcare utilization. This is to some extent surprising because one of the typically stated reasons for having early retirement is to provide an opportunity to withdraw from the labor force before employment leads to severe effects on health. In this section, we examine whether the average treatment effect in Section 4 masks substantial effects on vulnerable subpopulations. We do this by characterizing the different reform compliance strata in Section 5.1, estimating heterogeneous treatment effects using causal forest in Section 5.2, and lastly, estimating different treatment effect parameters in Section 5.3.

5.1 Characterization of Compliers

To further understand the mechanisms underlying our results, we now characterize the groups of compliers (those who work due to the reform), always-takers (those who are in employment irrespective of the reform), and never-takers (those who are not employed irrespective of the reform) by their average observable characteristics.

As illustrated in Table 1 and Figure 4, not all individuals were in employment in response to the reform-induced ERA increase. Individuals who are not working may rely on public transfers other than retirement benefits or use their savings to self-finance early retirement by becoming self-supporting. We find support for this in Section 4.1. In this section, we characterize the reform compliers by using exposure to the reform, $\mathbb{1}[X_i \geq c]$, as an instrument for employment, $D_i(\mathbb{1}[X_i \geq c])$, where $D_i(\mathbb{1}[X_i \geq c])$ denotes individual i's potential treatment status (employment) as a function of the instrument. As in Marbach and Hangartner (2020), we need to assume monotonicity (i.e., $D_i(1) \geq D_i(0)$) and independence of the instrument. Independence of the instrument implies $D_i(1), D_i(0), \mathcal{X}_i \perp \mathbb{1}[X_i \geq c]$, where $\mathcal{X}_i$ denotes the set of covariates. Hence, independence of the instrument states that instrument assignment is independent of an individual's compliance stratum and covariate value. For $\mathcal{X}_i$, we use the following covariates measured prior to the announcement: Gender, employment age 55, married age 55, dummy for high education age 55, average disposable income aged 48-42, average net wealth aged 48-42, GP visists aged 52-54, and the following health-related outcomes measured aged 47-54: Painkillers DDDs, antidepressant DDDs, CVD medicine DDDs, the Charlson Comorbidity Index (CCI), stroke (0/1), days hospitalized, and number of hospital contacts. Lastly, for the subgroup that was employed at age 55, we study the following occupational dummies measured at age 55: white-collar occupation, high physical activity, high routine work, high decision-making freedom, cramped workspace, high competitive pressure, hazardous work conditions, and high codifiability of work.[24]

The results are depicted in Figure 8. The share of compliers in Figure 8a corresponds to the reform's effect on employment in Column (1) of Table 1. Figure 8a shows a modest share of never-takers, and that compliers consist of a larger share of women and married individuals compared to the full population. Figures 8b to 8f highlight several key aspects of who the increased ERA targeted. Figure 8b shows that while compliers generally have lower education, they are also characterized by much lower levels of disposable income and net wealth, which suggests that the increase in ERA mostly targets financially vulnerable individuals. Figures 8c and 8d show that never-takers are characterized by worse pre-reform health and healthcare utilization outcomes compared to compliers and always-takers, measured by both GP visits, painkillers, antidepressants, and CVD medicine (Figure 8c) as well as more severe health-related outcomes such as CCI, stroke incidence, days hospitalized, and hospital contacts (Figure 8d). Overall, compliers are in worse health than always-takers, but are more similar to always-takers than to never-takers in terms of predetermined health and healthcare utilization, albeit with considerable uncertainty. This is a first indication that stricter reforms, which will push more people into employment, and therefore people from the never-takers into the complier group, will potentially have larger health effects, since this group is already in bad health. For individuals employed at age 55, Figures 8e and 8f show that always-takers differ systematically from compliers and never-takers in their occupational characteristics. Specifically, always-takers are more likely to hold white-collar jobs, work in roles with lower physical demands, less routine work, more decision-making freedom (Figure 8e), are less likely to work in jobs with cramped work environments, hazardous work conditions, codifiable work, and more often have decision-making autonomy. The opposite holds for the compliers for almost all occupational characteristics, whereas the never-takers are between always-takers and compliers.

The characterization of compliance strata shows that individuals who are employed because of the increased ERA have very distinct financial, predetermined health and healthcare utilization, and occupational characteristics. In Section 5.2, we examine whether these characteristics are associated with larger treatment effects by estimating heterogeneous treatment effects using causal forest. Under the assumption that individuals of worse pre-reform health also would suffer the largest deterioration in health,

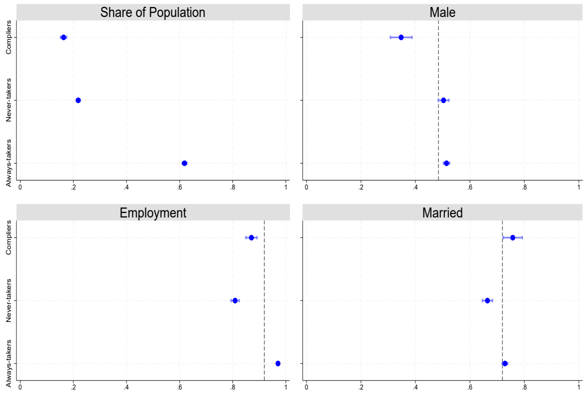
[24] We construct binary indicators for high physical activity, high routine work, high decision-making freedom, cramped workspace, high competitive pressure, hazardous work conditions, and high codifiability of work using $O^{*}NET$ (Occupational Information Network, developed under sponsorship of U.S. Department of Labor) data matched to Danish occupation codes. An indicator equals 1 if the corresponding measure is at or above the median of the overall sample of 55-year-olds.

this suggests that a more “strict” implementation of the 2011 reform,[25] inducing more individuals into employment would yield larger health and healthcare effects. We examine this point in Section 5.3.

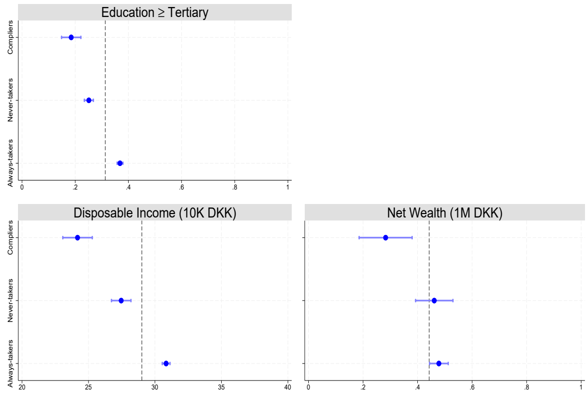
[25] A more ”strict” implementation of the 2011 reform could take various forms: For example, simultaneously tightening eligibility for UI benefits, reducing the early retirement benefits levels (thereby lowering lifetime income), or restricting access to pension wealth before retirement.

Figure 8: Characterization of ERA Compliers

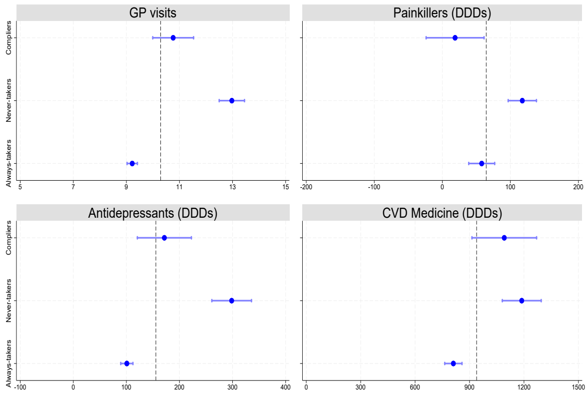
(a) Sample Share, Gender, Empl., and Marital Status



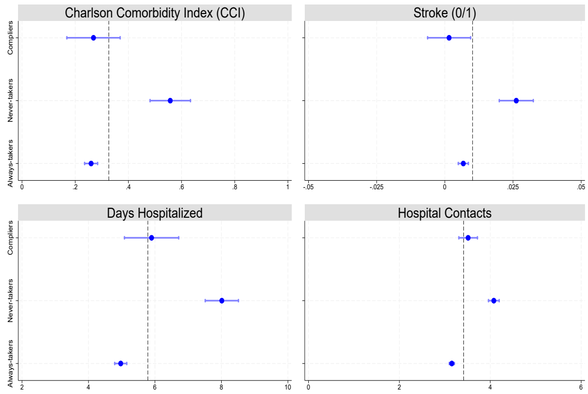
(b) Educational Level, Income, and Net Wealth



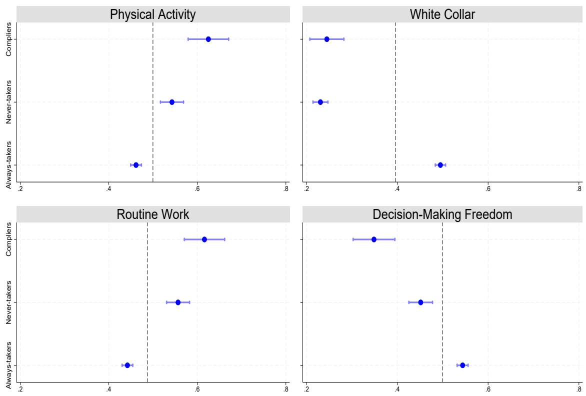
(c) GP Visits and Prescription Drugs (DDDs)



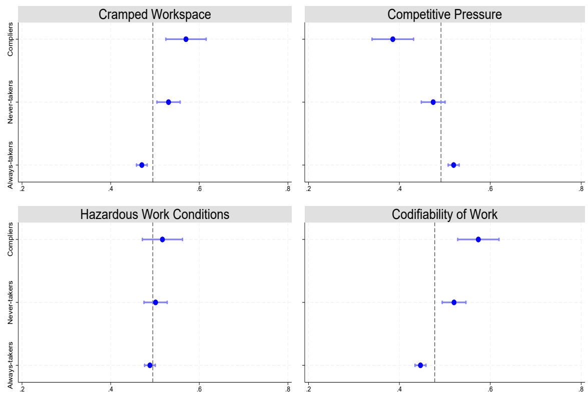
(d) CCI, Stroke, and Hospital Days and Contacts



(e) Occupation: Physical, White Collar, Routine, and Freedom



(f) Occupation: Cramped, Competitive, Hazardous, and Codifiability



The figure characterizes the compliance strata of compliers, always-takers, and never-takers as described in Section 5.1 for the ERA reform. The dashed, vertical lines indicate the full sample mean of each predetermined covariate. Panels (e)-(f) only include individuals employed at age 55 for which occupational characteristics can be measured. 95% confidence intervals are based on standard errors calculated asymptotically.

Different retirement reforms likely target different populations, whose responses may differ substantially. To illustrate this, we use that the reform also raised the statutory retirement age (SRA) in steps from 65 in 2019 to 67 in 2022, as shown in Figure 2. We repeat the complier characterization for the SRA increase in Appendix D.3.[26] Figure D.4 shows that, compared to the ERA increase, the share of never-takers is much higher (around 75%) and the share of compliers lower, resulting in wider confidence intervals for the compliers. Overall, the never-takers are in worse health than always-takers, consistent with what we find in the complier characterization of the ERA increase. However, as Figures D.4c and D.4d show, never-takers of the SRA increase resemble the overall population far more than the never-takers do under the ERA increase and compliers resemble more always-takers. Figure D.4 illustrates clearly that changes to different retirement schemes can target vastly different populations. The fact that even apparently similar changes in retirement ages target very different populations is of high importance when comparing findings across studies with different settings.

5.2 Treatment Effect Heterogeneity using Causal Forest

Examining the potential existence of heterogeneous treatment effects is important for at least three reasons. First, the main estimation results presented in Section 4 may mask significant heterogeneity across individuals. Unmasking this heterogeneity is informative about the underlying mechanisms driving the main results. Second, treatment heterogeneity can guide future retirement policies by identifying subpopulations that are less responsive in terms of healthcare utilization. Third, our complier analysis suggests that different reforms may impact some segments of the population more than others. Heterogeneous treatment effects may explain the lack of consensus across studies that use different retirement reforms for identification as these reforms possibly apply to distinct populations (e.g., male construction workers in Bauer and Eichenberger (2021) or female local government workers in Hagen (2018)). In contrast, our analysis includes both men and women, and although selection into the early retirement scheme is endogenous, individuals with both high and low educational attainment are represented in the sample (see also Table B.1). Understanding the heterogeneous treatment effects may thus increase the external validity of our findings.

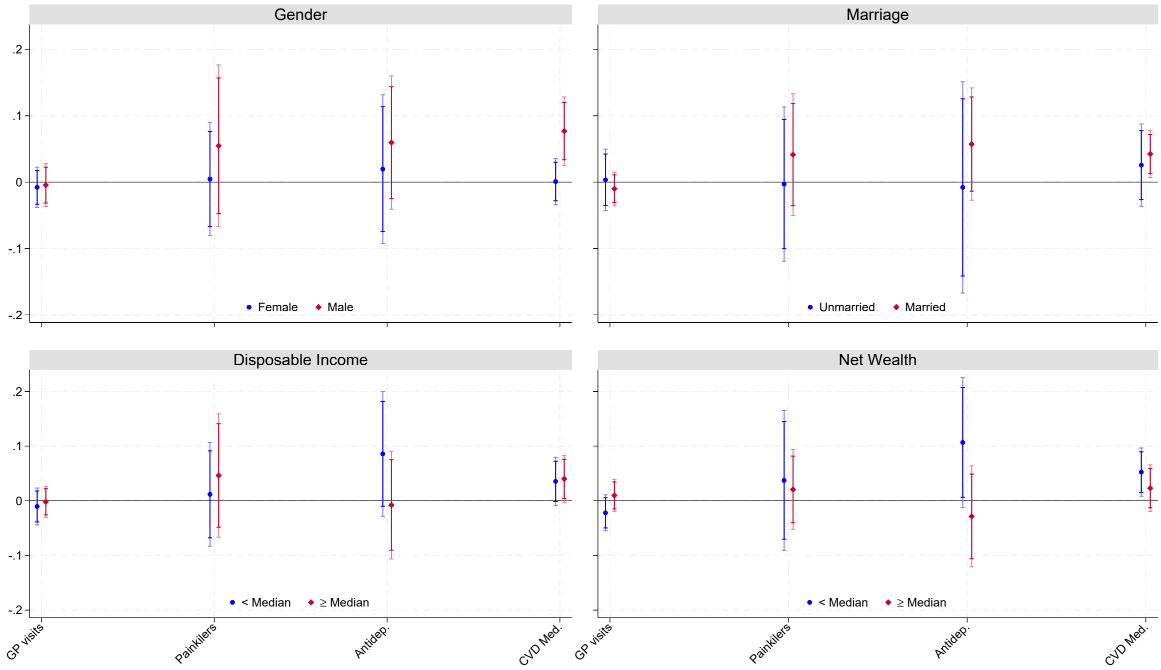
We estimate Conditional Average Treatment Effects (CATEs) for each individual based on 41 variables (see Appendix E for details) on individual, occupational, and workplace characteristics using Causal Forest following Athey et al. (2019) and Mayer et al. (2020). Figures 9 and 10 illustrate the group average treatment effects (GATEs) for 32 selected subgroups. We use doubly robust estima-

[26]Note, since we condition our sample on having paid contributions to the ERA scheme before the reform announcement, we do not target the full population affected by the SRA increase.

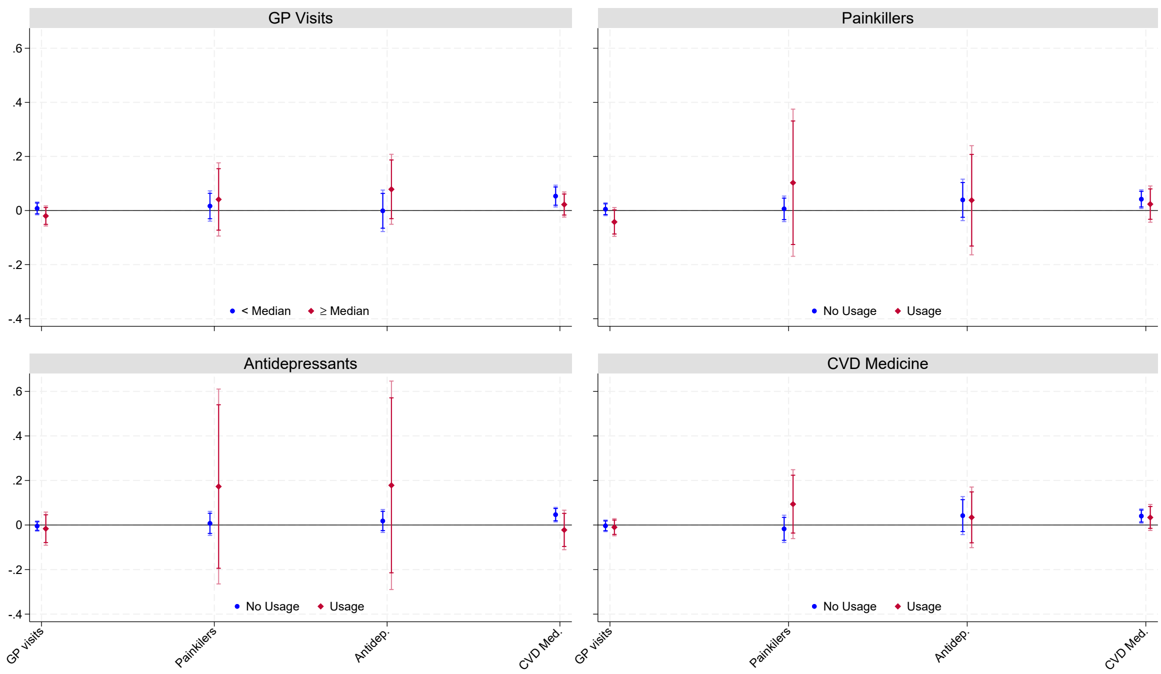
tors and augmented inverse-propensity weighting (AIPW) to estimate GATEs and to perform inference. When interpreting the GATEs, it is important to recognize that different groups consist of different compliance stratums, as evident in Section 5.1. If extending employment because of the reform is a significant treatment effect mediator, then the groups that consist of more compliers are also likely to exhibit higher treatment effects. Most GATEs in Figures 9 and 10 are not statistically significant at the 10% level. The effects on antidepressants and CVD medicine appear to have the highest variability across groups.

Figure 9: Estimated Group Average Treatment Effects (1/2)

(a) Gender, Marital Status, Income, and Net Weatlth



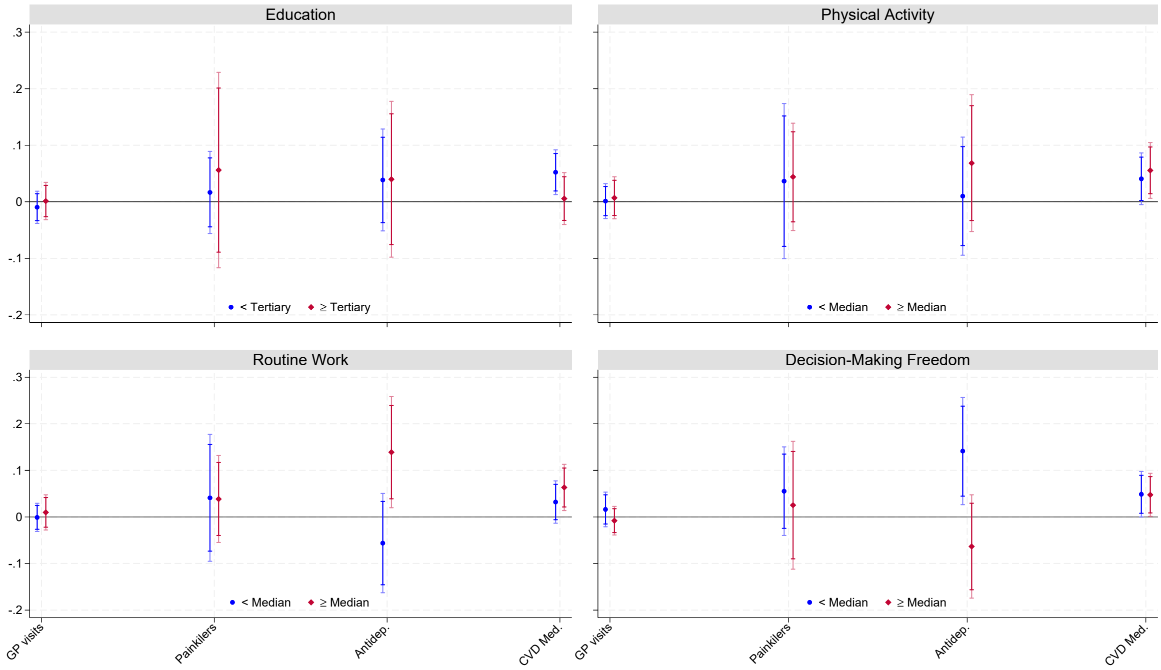
(b) Previous Healthcare Utilization



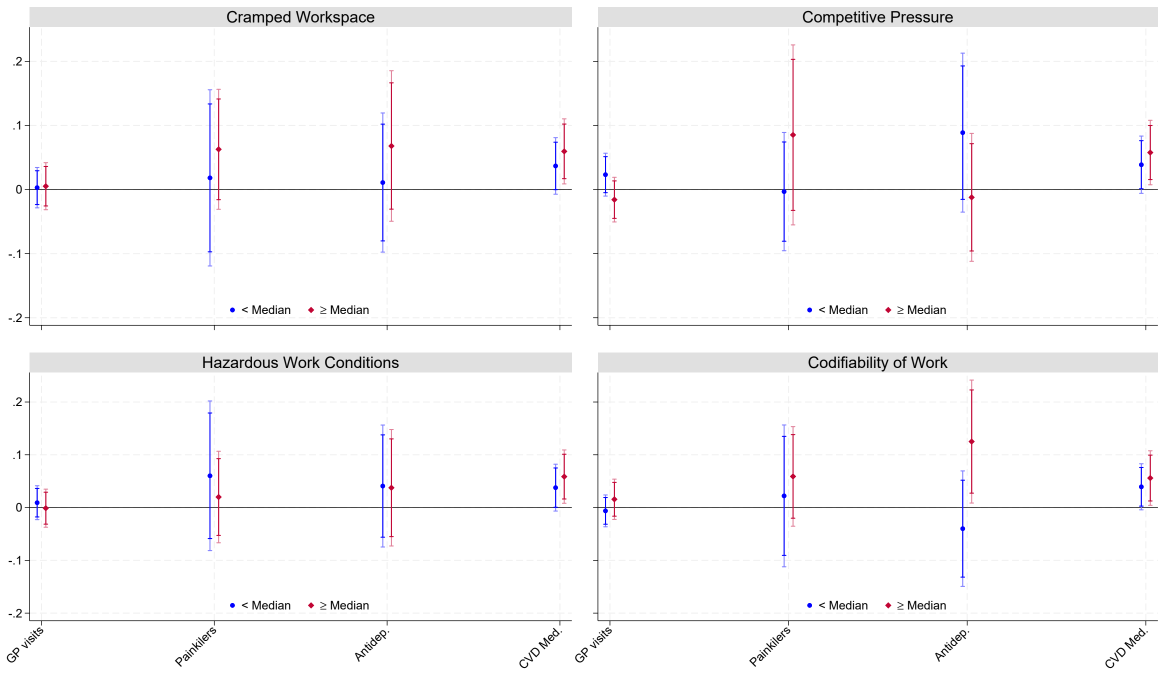
The figure shows the Group Average Treatment Effects (GATEs) for our four main outcomes estimated using causal forests as described in Section 5.2. All characteristics are measured before the announcement of the retirement reform, and outcomes indicated on the horizontal axes are measured 2 years after the implementation of the increased ERA. All effects are rescaled by the predicted outcome absent the reform estimated by random forests. The figure shows 95% and 90% confidence intervals. GATEs and inference are estimated using doubly robust estimators via augmented inverse-propensity weighting (AIPW).

Figure 10: Estimated Group Average Treatment Effects (2/2)

(a) Education, and Occupation: Physical, Routine, and Freedom



(b) Occupation: Cramped, Competition, Hazardous, and Codifiability



The figure shows the Group Average Treatment Effects (GATEs) for our four main outcomes estimated using causal forests as described in Section 5.2. All characteristics are measured before the announcement of the retirement reform, and outcomes indicated on the horizontal axes are measured 2 years after the implementation of the increased ERA. All effects are rescaled by the predicted outcome absent the reform estimated by random forests. The figure shows 95% and 90% confidence intervals. GATEs and inference are estimated using doubly robust estimators via augmented inverse-propensity weighting (AIPW).

Figure 9a shows that overall, males and the married are more responsive in terms of painkillers, antidepressants, and CVD medicine. Individuals with relatively low disposable income or net wealth also appear more responsive in terms of antidepressants. The effects on CVD medication for males, the married, and those with low net wealth are statistically significant, as is the effect on antidepressants for individuals with low net wealth. Figure 9b depicts a much larger estimated GATE on painkillers for individuals with prior usage of painkillers, and similarly, the estimated GATE on antidepressants is much larger for individuals with prior usage of antidepressants. However, there are also relatively few individuals with prior usage of these prescription drugs, as illustrated by the wide confidence intervals. For the use of CVD medicine, the estimated GATE is relatively similar, irrespective of prior use of CVD medicine. Figures 10a and 10b illustrates that occupational characteristics may be an important determinant for the magnitude of the treatment effect, especially for the use of antidepressants. In particular, the GATEs on antidepressants are much larger for groups whose occupations are characterized by relatively high degrees of routine work codifiability,[27] and a low degree of decision-making freedom. The GATEs for these individuals are also significant at the 5% level. Occupations that consist of relatively high routine, highly codifiable tasks with low decision-making freedom may be thought of as more "mundane" occupations for which it is plausible that individuals working in these are more severely affected in terms of antidepressants when extending employment. For the GATEs on CVD medicine, the largest treatment effect heterogeneity is across education levels, where those with relatively low education exhibit the highest treatment effect, being significant at the 5% level. While differences in treatment effects also occur for CVD medicine across different occupational types, these differences are much smaller.

Overall, Figures 9 and 10 illustrate that modest treatment effect heterogeneity is present for antidepressants and CVD medicine, with treatment effects being much more homogeneous for GP visits and painkillers. Males, and those with relatively low income or net wealth may have been particularly affected by the ERA reform. Those with relatively "mundane" (high routine and codifibabiltiy, low decision-making freedom) occupations were also affected more by the ERA reform in terms of antidepressants. While individuals with previous usage of prescription drugs may also have been more affected by the ERA reform, these effects are associated with high uncertainty, and we find no significant effects.[28]

[27]Codifiability refers to "*procedural, rule-based activities to which computers are currently well-suited*" (Acemoglu and Autor, 2011).

[28]Again, it should, however, be noted that treatment effect heterogeneity appears because of compliance strata differing between groups. For instance, we also find that those working in occupations with low decision-making freedom (see Figure 8f) were also more likely to extend employment because of the reform.

5.3 Marginal Health and Healthcare Utilization Effects of Increased ERA

The characterization of the never-takers in Section 5.1 shows that the never-takers of the extended ERA significantly differ from the compliers in many observable characteristics. The effects on health and healthcare utilization of "stricter" implementations of retirement reforms,[29] which push individuals from the never-taker group into the complier group, could therefore lead to larger effects on health and healthcare utilization. By treating exposure to the retirement reform as-if randomly assigned in a bandwidth around the cutoff and estimating the probability of being employed in the half-year following the ERA extension, we then estimate marginal treatment effects (MTEs) by tracing out the effects of increased employment along different margins of the unobserved resistance to employment. Following the same principles and data selection as in Section 4.2, we estimate MTEs that are "local" to individuals born around the cutoff given by $\mathcal{W}$. The MTE framework allows us to assess the extent and pattern of treatment effect heterogeneity by both observed and unobserved characteristics.

To estimate MTEs, we use the Local Instrument Variable (LIV) approach introduced by Björklund and Moffitt (1987) and generalized by Heckman and Vytlacil (2005, 2007). The traditional LIV approach requires a continuous instrument, exploiting that one can then compare any pair of values, z and $z^{'}$ of the instrument, Z_i, and calculate covariate-specific Wald estimators for these values (Cornelissen et al., 2016). As our instrument is binary, i.e., exposure/non-exposure to the increased ERA as determined by date of birth, we instead trace out a distribution of propensity scores with continuous support by combining our instrument, $\mathbb{1}[X_i \geq c]$, with several (continuous) covariates, $\mathcal{X}_i$. Similar strategies, using continuous covariates to trace out a distribution of propensity scores, are followed by, e.g., Kline and Walters (2016), Brinch et al. (2017), Walters (2018), Cornelissen et al. (2018), and Agan et al. (2023).

We describe our approach for estimating MTEs in detail in Appendix F.[30] It is worth noticing that estimating MTEs requires the usual IV assumptions, relevancy and the exclusion restriction of the instruments, and, additionally, the separability assumption. While relevancy is implicitly shown from Column (1) of Table 1, we have not imposed the exclusion restriction anywhere else so far. As argued in Section 3, the exclusion restriction is not without bite, as it rules out cases where the reform itself, rather than extended employment, could affect health and healthcare utilization. The exclusion restriction therefores rules out e.g., increased depression risk due to perceptions of unfair treatment by the reform. Last, separability is a strong assumption as well, but it is needed to identify the MTE over the common support of the propensity score, unconditional on observable characteristics. Intuitively, separability imposes

[29] A "stricter" implementation could be e.g., simultaneously making it harder to obtain other transfer types or by limiting their payouts, or alternatively making it more difficult to withdraw pension wealth.

[30] In practice, we estimate the MTEs and treatment parameters by using the `mtefe` Stata-package by Andresen (2018).

that selection on unobservables works "the same way" for every value of the observable characteristics (Kline and Walters, 2016). Consequently, while the MTE framework allows, e.g., individuals of high and low education to have different health and healthcare utilization effects of extending employment, separability imposes that if low education individuals are selected into extending employment based on these health and healthcare utilization gains, then high education individuals must as well.

In the main specification used in Table 4, we assume that the MTE curves take the form of third-order polynomials. In Table F.2 and Figure F.2 we show that the overall findings are relatively robust to increasing the flexibility of the functional form. It should also be noted that we trim 1% of the sample based on the highest and lowest propensity scores similarly to Carneiro et al. (2011) to alleviate issues with common support. As a consequence, the estimated treatment effect parameters are only identified for the trimmed sample.

5.3.1 Estimated MTEs

The results of estimating MTEs as described in Appendix F are depicted in Table 4 and Figure 11. We report four treatment effect parameters: The Average Treatment Effect (ATE), Local Average Treatment Effect (LATE), Average Treatment Effect on the Treated (ATT) and the Average Treatment Effect on the Untreated (ATUT). In Table F.2 we also report the marginal policy-relevant treatment effect (MPRTE), defined as the treatment effect of a policy intervention that increases all propensity scores by an infinitesimally small fraction (see Appendix F). The ATEs are a weighted average of the ATTs and ATUTs, while the LATEs correspond roughly to the treatment effects reported in Table 2, scaled by the first stage in Table 1.[31] It should be noted that most of the estimated treatment effect parameters in Table 4 are not statistically significant. Despite this, the results highlight several interesting findings.

The only ATEs that are statistically significant at the 10% level are for GP visits (Column (1)) and CVD medicine (Column (4)). For GP visits, antidepressants, and CVD medicine (Columns (1) and (3)-(4)), the estimated average treatment effects on the treated (ATTs) are much lower than the estimated average treatment effects on the untreated (ATUTs). This is a result of the generally upward sloping MTE curves illustrated in Figures 11a, 11c and 11d, although none of the slope coefficients are statistically significant.[32] Upward-sloping MTE curves are consistent with reverse selection on gains. Reverse selection on gains suggests that either individuals or employers act as if they possess some

[31] Discrepancies arise because we include a larger set of observable variables in Table 4, include interactions between the instrument and observables, and weightings differ.

[32] See the row "P-value, k(p)" in Table 4. The slope coefficients for CVD medicine are borderline insignificant at the 10 % level.

Table 4: Estimated Treatment Effect Parameters

	(1) GP visits	(2) Painkillers DDDs	(3) Antidepressants DDDs	(4) CVD Medicine DDDs
LATE	-0.0797	-9.706	15.26	130.3*
	(0.441)	(12.17)	(13.83)	(66.73)
ATE	-1.312*	9.756	30.79	172.7**
	(0.679)	(22.20)	(26.45)	(87.85)
ATT	-2.229**	23.85	22.69	120.5
	(1.095)	(37.60)	(44.39)	(137.8)
ATUT	0.700	-21.42	48.89*	285.8**
	(0.819)	(24.95)	(27.25)	(124.5)
P-value, k(p)	0.242	0.424	0.752	0.101
P-value, $\beta_1 - \beta_0$	0.009	0.686	0.003	0.006
Controls	Yes	Yes	Yes	Yes
N	23,456	23,456	23,456	23,456

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors in parentheses with 1,000 replications.

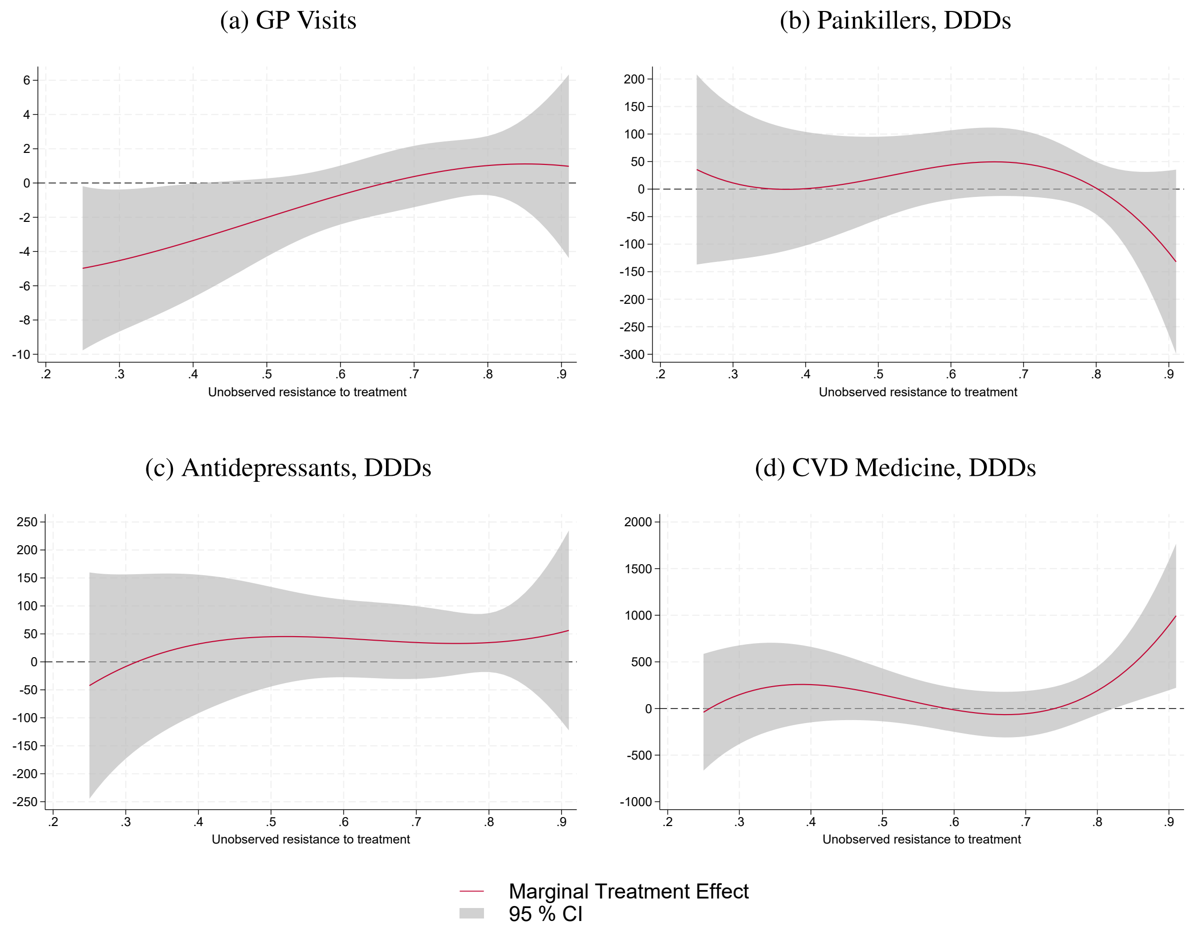
The table shows treatment effect parameters estimated as described in Appendix F. All outcomes denoted in the column headers are measured two years after implementation of the extended ERA. "Controls" included in the estimations are average wealth and disposable incomes at ages 48-52, including squared terms, and dummies for gender, being married, and education level, as well as healthcare utilization measured four years before announcement of the reform, and occupational characteristics measured at age 55. The excluded instruments are a binary variable for being born after July 1, 1954, and this binary variable interacted with the variables for disposable income and net wealth and their squared terms. The row "P-value, k(p)" refers to the P-value from a joint test of the slope coefficients of the MTE equation, Equation (F5), and "P-value, $\beta_1 - \beta_0$" refers to the P-value from a joint test of whether the treatment effects differ across observed covariates.

knowledge of the idiosyncratic effect of increasing employment on health and healthcare utilization, and base their employment decision on this. E.g., reverse selection on gains could arise if the most worn-out workers choose to quit their jobs despite being affected by the reform, because they know extending employment would otherwise lead to an increase in CVD medicine consumption. In contrast, for painkillers, the estimated ATT is much larger than the estimated ATUT. We also find evidence of treatment effect heterogeneity by observables for GP visits, antidepressants, and CVD medicine,[33] which is consistent with our findings Section 5.2, with the exception of GP visits, for which we found limited heterogeneity.

Taken together, the estimated treatment effect parameters in Table 4 suggest that implementing retirement reforms that induce larger increases in employment may have larger effects on GP visits, an-

[33] See the row "P-value, $\beta_1 - \beta_0$" in Table 4.

Figure 11: Estimated Marginal Treatment Effects



The figure illustrates the MTEs estimated from Equation (F5) in Appendix F, where $k(u)$ is a polynomial of order $L = 3$, and x is evaluated at mean levels. The 95 % confidence intervals given by gray areas are calculated using bootstrapped standard errors with 1,000 replications.

tidepressants, and CVD medicine, and smaller effects on painkillers, compared to our main results. It should be noted that almost all of the estimated treatment effect parameters are associated with significant uncertainty. These findings caution policymakers against interpreting our main results as small and borderline (in)significant for *any* type of retirement reform. Our main estimates may not generalize to "stricter" retirement reforms, which could have substantially larger effects on health and healthcare utilization.

6 Discussion and Conclusion

The increasing ratio of retirees to active workers is putting significant financial pressure on the pay-as-you-go social security systems of most developed countries (OECD, 2023). Retirement reforms aimed at encouraging older workers to work for longer have been and are still being widely adopted to relieve this financial pressure. A concern often raised in political debates about this subject is whether the affected workers, who are on the margin of retirement, suffer adverse health effects from delayed retirement. Consequently, a *key* policy concern is to design and implement retirement policies that increase the employment of older workers while minimizing potential adverse health effects.

In this paper, we explore the relationship between a reform that increased the retirement age and health and healthcare utilization. We do this by using the 2011 retirement reform that increased the early retirement age (ERA) in a regression discontinuity design applied to detailed Danish administrative data. Our findings offer important insights into the multifaceted consequences of retirement reforms that aim to extend the employment of older workers by raising the age of eligibility for publicly provided pension schemes. The key contributions of this study are threefold and have significant policy implications.

First, we establish that the incremental half-year increase in the early retirement age not only raised employment for the targeted age group but also led to a notable increase in the uptake of other public transfers and the number of self-supporting individuals (i.e., those who are neither employed nor receiving public transfers). This demonstrates the interplay between retirement policies and broader social security and pension systems. The group of individuals that increased employment in response to the reform was characterized by lower disposable income and net wealth pre-reform, indicating that retirement policies disproportionately affect those with fewer economic resources.

Second, we find no effect of an increased ERA on GP visits and usage of painkillers. For antidepressants and CVD medicine, we find small positive effects that are borderline significant, though these effects are sensitive to model specification. This suggests that well-anticipated reforms with a staggered rollout that increase the early retirement age (ERA) from a relatively low level (beginning of 60s) could possibly increase employment among older individuals with negligible health effects. While we find no apparent trade-off between increasing the employment of older workers and health or healthcare utilization for the overall population, our results indicate that some subgroups may experience adverse health effects. Using causal forest analysis, we show evidence of heterogeneous treatment effects, e.g., men significantly increasing CVD medicine usage following the ERA increase, while women are unaffected. Our results also indicate that individuals in highly routine occupations with limited decision-making freedom are more likely to increase their antidepressant use in response to delayed retirement. Hence, our results suggest that for some subgroups, increasing the ERA might still have adverse effects on health and healthcare utilization, but overall health-related effects are limited.

Third, our study highlights the role of contextual factors in designing future retirement reforms. Individuals may claim early retirement either because they experience a high disutility from work (e.g., poor health) or because they place a high value on leisure. Policy discussions have focused on which of these groups is most affected by the reform, and, consequently, whether increased employment might come at the cost of greater health problems. We show that never-takers of the reform, i.e., those who did not work despite the policy change, were characterized by worse health outcomes before the reform, while compliers (individuals whose employment in the half-year following the ERA increase occurs *due* to the increase in the ERA) in general were in better health than never-takers, but worse health than always-takers. This suggests that stricter retirement reforms, e.g., reducing retirement benefit levels and thereby lowering lifetime income, could lead to more pronounced health impacts, a hypothesis supported by our marginal treatment effect analysis. Taken together, we argue that our findings reveal that the outside options to employment (transfers or self-financed retirement), may serve as a mitigating channel, suggesting the importance of social security systems that can cushion the adverse effects of retirement reforms.

The overall limited effects of the increased ERA on healthcare utilization may be attributed to several factors. The gradual, stepwise implementation of the increased ERA allowed individuals to adapt. Because the reform was announced well in advance, individuals could incorporate it into their decisions, which we argue makes our estimates more reflective of long-run outcomes. Another factor could be the relatively low starting point of the ERA (60½ years for 1954:Q2-1954:Q3 cohort and 61½ years for 1955:Q2-1955:Q3 cohort), which may have attenuated effects. However, we do not find much larger effects for the younger 1956 cohort, whose ERA was higher. We can also only identify effects in the relatively short term, up to 6.5 years after the effect of the reform, meaning we cannot rule out longer-run effects. However, since we find at most limited and sensitive short-run health-related effects (generally negligible on average but with some heterogeneity across subgroups), it appears unlikely that sizable long-run health consequences would emerge.

Bibliography

3F (2019). Smerter: Hver anden 3F'er tager medicin på grund af jobbet. *Fagbladet 3F*.

Acemoglu, D. and Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

Agan, A., Doleac, J. L., and Harvey, A. (2023). Misdemeanor Prosecution. *The Quarterly Journal of Economics*, 138(3):1453–1505.

Ahammer, A. and Packham, A. (2023). Effects of unemployment insurance duration on mental and physical health. *Journal of Public Economics*, 226:104996.

Andresen, M. E. (2018). Exploring Marginal Treatment Effects: Flexible Estimation Using Stata. *The Stata Journal*, 18(1):118–158.

Arnold, G. and Depew, B. (2018). School starting age and long-run health in the United States. *Health Economics*, 27(12):1904–1920.

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

Athey, S., Friedberg, R., Mühlbach, N. S., Steimer, H., and Wager, S. (2020). *Between Work, Public Programs, and Retirement: Heterogeneous Responses to a Retirement Reform. Chapter 2 of PhD thesis by N. S. Mühlbach: Essays in Applied Econometrics and Causal Machine Learning.* PhD thesis, Aarhus University.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148 – 1178.

ATP (2024). Pension i tal (Pension - key figures). https://www.atp.dk/vores-opgaver/atp-livslang-pension/atps-pensionsprodukt/pensionsanalyser/pension-i-tal.

Bajari, P., Hong, H., Park, M., and Town, R. (2011). Regression Discontinuity Designs with an Endogenous Forcing Variable and an Application to Contracting in Health Care. Working Paper 17643, National Bureau of Economic Research.

Banks, J., Cribb, J., Emmerson, C., and Sturrock, D. (2025). The impact of work on cognition and physical disability: Evidence from English women. *Labour Economics*, page 102730.

Barreca, A. I., Guldi, M., Lindo, J. M., and Waddell, G. R. (2011). Saving Babies? Revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics*, 126(4):2117–2123.

Barreca, A. I., Lindo, J. M., and Waddell, G. R. (2016). Heaping-induced bias in regression-discontinuity designs. *Economic Inquiry*, 54(1):268–293.

Barschkett, M., Geyer, J., Haan, P., and Hammerschmid, A. (2022). The effects of an increase in the retirement age on health — Evidence from administrative data. *The Journal of the Economics of Ageing*, 23:100403.

Bauer, A. B. and Eichenberger, R. (2021). Worsening workers' health by lowering retirement age: The malign consequences of a benign reform. *The Journal of the Economics of Ageing*, 18:100296.

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

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

Björklund, A. and Moffitt, R. (1987). The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. *The Review of Economics and Statistics*, 69(1):42–49.

Blundell, R., French, E., and Tetlow, G. (2016). Retirement Incentives and Labor Supply. In Piggott, J. and Woodland, A., editors, *Handbook of the Economics of Population Aging*, volume 1, chapter 8, pages 457–566. Elsevier.

Bozio, A., Garrouste, C., and Perdrix, E. (2021). Impact of later retirement on mortality: Evidence from France. *Health Economics*, 30(5):1178–1199.

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

Brugiavini, A., Buia, R. E., Pasini, G., and Weber, G. (2026). The Effects of Pension Reforms on Health Inequality in Italy. Working Paper 34655, National Bureau of Economic Research.

Caliendo, M., Mahlstedt, R., van den Berg, G. J., and Vikström, J. (2023). Side effects of labor market policies. *The Scandinavian Journal of Economics*, 125(2):339–375.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Carneiro, P., Heckman, J. J., and Vytlacil, E. (2010). Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. *Econometrica*, 78(1):377–394.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–2781.

Carrino, L., Glaser, K., and Avendano, M. (2020). Later retirement, job strain, and health: Evidence from the new State Pension age in the United Kingdom. *Health Economics*, 29(8):891–912.

Cattaneo, M. D., Frandsen, B. R., and Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate. *Journal of Causal Inference*, 3(1):1–24.

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2023). A Practical Introduction to Regression Discontinuity Designs: Extensions. *arXiv preprint arXiv:2301.08958*.

Cattaneo, M. D. and Titiunik, R. (2022). Regression Discontinuity Designs. *Annual Review of Economics*, 14(1):821–851.

Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A., and Kerber, R. A. (2003). Association between telomere length in blood and mortality in people aged 60 years or older. *The Lancet*, 361(9355):393–395.

Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, 40(5):373–383.

Christensen, S., Johansen, M. B., Christiansen, C. F., Jensen, R., and Lemeshow, S. (2011). Comparison of Charlson comorbidity index with SAPS and APACHE scores for prediction of mortality following intensive care. *Clinical Epidemiology*, 3:203–211.

Ci, Z. (2022). Does raising retirement age lead to a healthier transition to retirement? Evidence from the U.S. Social Security Amendments of 1983. *Health Economics*, 31(10):2229–2243.

Contoyannis, P., Hurley, J., Grootendorst, P., Jeon, S.-H., and Tamblyn, R. (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.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2016). From LATE to MTE: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41:47–60.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy*, 126(6):2356–2409.

Cremers, J., Nielsen, T. H., and Ekstrøm, C. T. (2024). The causal effect of early retirement on medication use across sex and occupation: evidence from Danish administrative data. *European Journal of Health Economics*, 25:1517–1527.

Crossley, T. F. and Kennedy, S. (2002). The reliability of self-assessed health status. *Journal of Health Economics*, 21(4):643–658.

Cutler, D. M., Lleras-Muney, A., and Vogl, T. (2008). Socioeconomic Status and Health: Dimensions and Mechanisms. Working Paper 14333, National Bureau of Economic Research.

Danish Economic Councils (2021). Dansk Økonomi, forår 2021.

Danish Ministry of Immigration and Integration (2023). Endeligt svar på spørgsmål nr. 44 (alm. del) – mfu spørgsmål om ministeren vil redegøre for fødselsdatoer blandt flygtninge og migranter. https://www.ft.dk/samling/20222/almdel/uui/spm/44/svar/1937349/2673354/index.htm. UUI, Alm.del - 2022-23 (2. samling).

Dee, T. S. and Sievertsen, H. H. (2018). The gift of time? School starting age and mental health. *Health Economics*, 27(5):781–802.

Duggan, M., Singleton, P., and Song, J. (2007). Aching to retire? The rise in the full retirement age and its impact on the social security disability rolls. *Journal of Public Economics*, 91(7-8):1327–1350.

Epel, E. S., Blackburn, E. H., Lin, J., Dhabhar, F. S., Adler, N. E., Morrow, J. D., and Cawthon, R. M. (2004). Accelerated telomere shortening in response to life stress. *Proceedings of the National Academy of Sciences*, 101(49):17312–17315.

Eriksson, K. M., Westborg, C.-J., and Eliasson, M. C. E. (2006). A randomized trial of lifestyle intervention in primary healthcare for the modification of cardiovascular risk factors The Björknäs study. *Scandinavian Journal of Public Health*, 34(5):453–461.

Frimmel, W. and Pruckner, G. J. (2020). Retirement and healthcare utilization. *Journal of Public Economics*, 184:104146.

García-Miralles, E. and Leganza, J. M. (2024a). Joint retirement of couples: Evidence from discontinuities in Denmark. *Journal of Public Economics*, 230:105036.

García-Miralles, E. and Leganza, J. M. (2024b). Public Pensions and Private Savings. *American Economic Journal: Economic Policy*, 16(2):366–405.

Grip, A. D., Lindeboom, M., and Montizaan, R. (2012). Shattered Dreams: The Effects of Changing the Pension System Late In the Game. *The Economic Journal*, 122(559):1–25.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2):223–255.

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

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

Heckman, J. J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738.

Heckman, J. J. and Vytlacil, E. J. (2007). Chapter 71: Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. *Handbook of Econometrics*, 6, Part B:4875–5143.

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

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Janlert, U., Asplund, K., and Weinehall, L. (1992). Unemployment and Cardiovascular Risk Indicators Data from the MONICA Survey in Northern Sweden. *Scandinavian Journal of Social Medicine*, 20(1):14–18.

Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. *The Quarterly Journal of Economics*, 131(4):1795–1848.

Kuusi, T., Martikainen, P., and Valkonen, T. (2020). The influence of old-age retirement on health: Causal evidence from the Finnish register data. *The Journal of the Economics of Ageing*, 17:100257.

Leth-Petersen, S. and Skipper, N. (2014). Income and the Use of Prescription Drugs for Near Retirement Individuals. *Health Economics*, 23(3):314–331.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4):773–793.

Manoli, D. and Weber, A. (2016a). Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions. *American Economic Journal: Economic Policy*, 8(4):160–182.

Manoli, D. S. and Weber, A. (2016b). The Effects of the Early Retirement Age on Retirement Decisions. Working Paper 22561, National Bureau of Economic Research.

Marbach, M. and Hangartner, D. (2020). Profiling Compliers and Noncompliers for Instrumental-Variable Analysis. *Political Analysis*, 28(3):435–444.

Mastrobuoni, G. (2009). Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities. *Journal of Public Economics*, 93(11-12):1224–1233.

Mayer, I., Sverdrup, E., Gauss, T., Moyer, J.-D., Wager, S., and Josse, J. (2020). Doubly robust treatment effect estimation with missing attributes. *The Annals of Applied Statistics*, 14(3):1409–1431.

McEwen, B. S. (1998). Protective and Damaging Effects of Stress Mediators. *New England Journal of Medicine*, 338(3):171–179.

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

Nielsen, T. H. (2016). The Relationship Between Self-Rated Health and Hospital Records. *Health Economics*, 25(4):497–512.

OECD (2023). *Pensions at a Glance 2023: OECD and G20 Indicators*. OECD Publishing, Paris.

OECD (2024). *OECD Pensions Outlook 2024: Improving Asset-backed Pensions for Better Retirement Outcomes and More Resilient Pension Systems*. OECD Publishing, Paris.

OECD (2024). Population statistics 2024. *OECD Publishing*.

OECD (2025). Benefits in unemployment, share of previous income. https://www.oecd.org/en/data/indicators/benefits-in-unemployment-share-of-previous-income.html.

OECD (2025). Government at a glance. *OECD Publishing*.

Ollonqvist, J., Kotakorpi, K., Laaksonen, M., Martikainen, P., Pirttilä, J., and Tarkiainen, L. (2025). Incentives, Health, and Retirement: Evidence from a Finnish Pension Reform. *Health Economics*, 34(3):537–572.

Pierce, J. R. and Schott, P. K. (2020). Trade Liberalization and Mortality: Evidence from US Counties. *American Economic Review: Insights*, 2(1):47–64.

Seeman, T. E., McEwen, B. S., Rowe, J. W., and Singer, B. H. (2001). Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. *Proceedings of the National Academy of Sciences*, 98(8):4770–4775.

Serrano-Alarcón, M., Ardito, C., Leombruni, R., Kentikelenis, A., d'Errico, A., Odone, A., Costa, G., Stuckler, D., and IWGRH (2023). Health and labor market effects of an unanticipated rise in retirement age. Evidence from the 2012 Italian pension reform. *Health Economics*, 32(12):2745–2767.

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

Simonsen, M., Skipper, L., and Skipper, N. (2016). Price Sensitivity of Demand for Prescription Drugs: Exploiting a Regression Kink Design. *Journal of Applied Econometrics*, 31(2):320–337.

Staubli, S. and Zweimüller, J. (2013). Does raising the early retirement age increase employment of older workers? *Journal of Public Economics*, 108:17–32.

Stuart, E. A. and Rubin, D. B. (2008). *Best practices in quasi-experimental designs*. SAGE Publications, Inc.

The Danish Clinical Quality Program (2024). Dansk Stroke Register: National Årsrapport 2023, 1. januar - 31. december 2023. Technical report, The Danish Clinical Quality Program– National Clinical Registries (RKKP).

Vestad, O. L. (2013). Labour supply effects of early retirement provision. *Labour Economics*, 25:98–109.

Vytlacil, E. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica*, 70(1):331–341.

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Walters, C. R. (2018). The Demand for Effective Charter Schools. *Journal of Political Economy*, 126(6):2179–2223.

A Overview of Selected Literature

Table A.1: Literature Overview

Study	Country	Data	Population	Reform	Implementation Lag	Empirical Strategy	Results	UI Generosity*
Panel A: Negative effect of increasing retirement age on health								
Grip et al. (2012)	NL	Survey, admin	Male public sector workers	Reduction in pension rights: Work 1 year and 1 month longer to achieve same replacement rate as pre-reform	6 months	OLS	Negative effect on self-reported mental health	64%
Shai (2018)	IL	Survey	Men	Increase in full retirement age (65 to 67)	3 months	DiD	Negative effect on health status and increased healthcare spending	20%
Barschkett et al. (2022)	DE	Admin	Women	Increase in early retirement age (60 to 63)	13 years	DiD	Negative effects on physical and mental health	60%
Carrino et al. (2020)	GB	Survey	Women	Gradual increase in state pension age (60 to 66)	15 years	DiD	Negative effect on self-reported mental health	17%
Serrano-Alarcón et al. (2023)	IT	Survey, admin	Women	Increase in normal retirement age (60 to 63 years gradually up to 67)	Less than 1 month	DiD	Negative effect on mental health (more hospitalizations) and overall health (more injuries)	61%
Panel B: No effect of increasing retirement age on health								
Hernaes et al. (2013)**	NO	Admin	Men, women	Gradual decrease in early retirement age (67 to 62)	Less than 1 year	IV, hazard rate model	No effect on mortality	67%
Hagen (2018)	SE	Admin	Female local government workers	Increase in normal retirement age (63 to 65)	1.5 years	DiD	No effect on mortality, hospitalizations, or prescription drugs	64%
Bozio et al. (2021)	FR	Admin	Private sector workers	Progressively increased contribution length for full pension by cohort	1 year	IV	No effect on mortality	66%
Panel C: Positive effect of increasing retirement age on health								
Bertoni et al. (2018)	IT	Survey	Men	Increase in minimum eligibility age (57 to 60 or 58 to 61)	3.5 years	DiD	Positive effects on health investments (more exercise)	61%
Bauer and Eichenberger (2021)**	CH	Survey	Male construction workers	Gradual decrease in retirement age (65 to 60)	Less than 1 year	DiD	Positive effect on self-reported health	72%
Ci (2022)	US	Survey	Men, women	Gradual increase in retirement age (65 to 67)	17 years	DiD	Positive effects on health (lower drug use, lower BMI)	9%

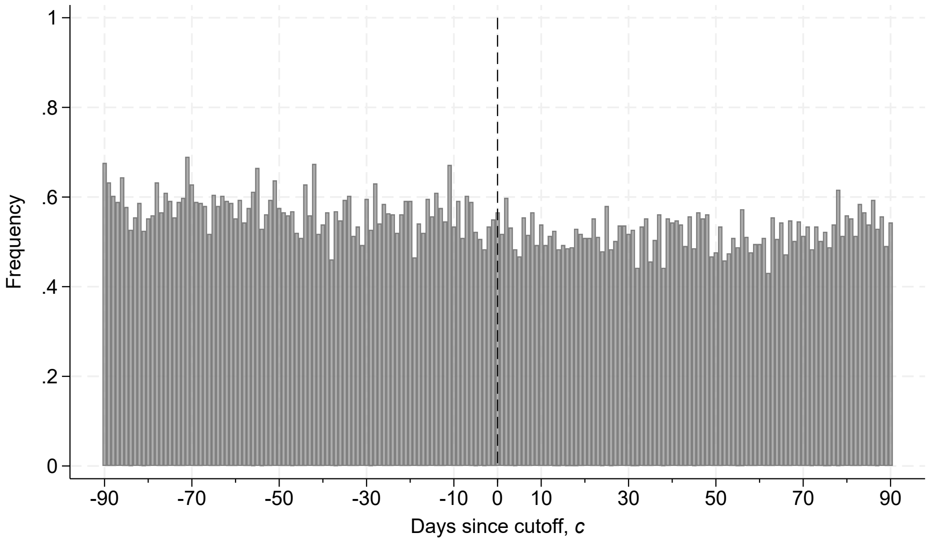
* UI generosity is proxied by the unemployment insurance (UI) net replacement rate, defined as benefits in unemployment divided by disposable household income before job loss for a single person with no children after 12 months in 2024. Source: OECD (2025)

** Assuming symmetric effects of decreasing retirement ages to increasing retirement ages.

B Details on Data and Variables

B.1 Additional Descriptive Figures and Tables

Figure B.1: Frequency of Birth Dates



The figure shows the frequency of birth dates in percent for 90 days on each side of the cutoff, *c* (July 1).

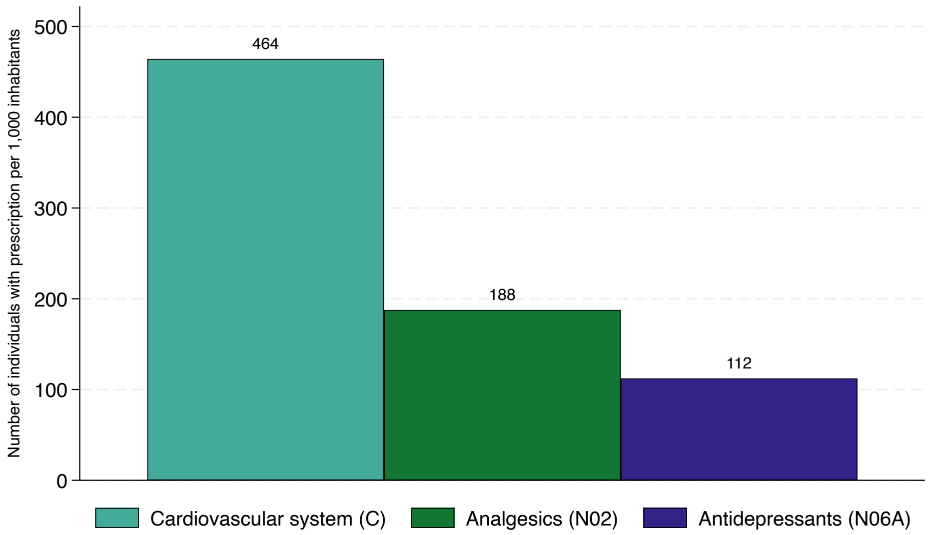
Table B.1: Sample Descriptives

	(1) Population		(2) Danish Origin		(3) Selected sample	
	Mean	Sd.	Mean	Sd.	Mean	Sd.
Gender and marital status:						
Male	0.50	0.50	0.50	0.50	0.48	0.50
Married	0.67	0.47	0.66	0.47	0.72	0.45
Education (shares):						
Primary, lower secondary, unknown	0.30	0.46	0.30	0.46	0.26	0.44
Upper secondary	0.40	0.49	0.40	0.49	0.42	0.49
Short cycle tertiary	0.05	0.21	0.05	0.21	0.05	0.21
Bachelor	0.18	0.39	0.18	0.39	0.20	0.40
Master or doctoral	0.07	0.26	0.07	0.26	0.07	0.25
Income and labor market outcomes:						
Disposable Income (10K DKK)	27.77	17.13	28.29	17.04	29.01	13.83
Net Wealth (1M DKK)	0.41	1.42	0.43	1.44	0.44	1.25
ERA payment, share	0.67	0.47	0.69	0.46	1.00	0.00
Employment, share	0.80	0.40	0.81	0.39	0.92	0.27
Health and healthcare utilization:						
GP visits	11.44	12.43	11.03	11.96	10.29	9.99
Painkillers (DDDs)	133.68	754.80	132.61	765.62	64.53	561.51
Antidepressants (DDDs)	228.38	862.67	228.48	870.62	155.52	654.15
CVD Medicine (DDDs)	1017.98	2505.20	1026.78	2523.99	939.00	2299.52
Painkillers Usage (0/1)	0.36	0.48	0.35	0.48	0.31	0.46
Antidepressants Usage (0/1)	0.20	0.40	0.20	0.40	0.16	0.37
CVD Medicine Usage (0/1)	0.47	0.50	0.47	0.50	0.46	0.50
CCI	0.45	1.73	0.44	1.70	0.33	1.28
Stroke (0/1)	0.01	0.12	0.01	0.12	0.01	0.10
Observations	36,814		34,263		23,694	

The table shows means and standard deviations for the population and the two sample restriction criteria noted in Section 2.3. Column (1) contains all individuals born within 50 days of July 1, 1954 or July 1, 1955. Column (2) is equivalent to Column (1), except excluding individuals who are not of "Danish origin", as defined by Statistics Denmark. Column (3) further excludes individuals who did not pay ERA contributions at age 55 and describes the main estimation sample. Marital status, education, and employment are measured at age 55. Income and net wealth (excluding pension wealth) are measured in DKK as the yearly average at ages 48-52, GP visits are measured as the total number of GP visits from age 52-54, and painkillers use, antidepressants use, CVD medicine use, CCI and incidence of stroke are measured over age 47-54. All monetary variables are deflated to DKK 2022 values using the CPI from Statistics Denmark. 1 DKK $\approx$ 0.1416 USD in 2022.

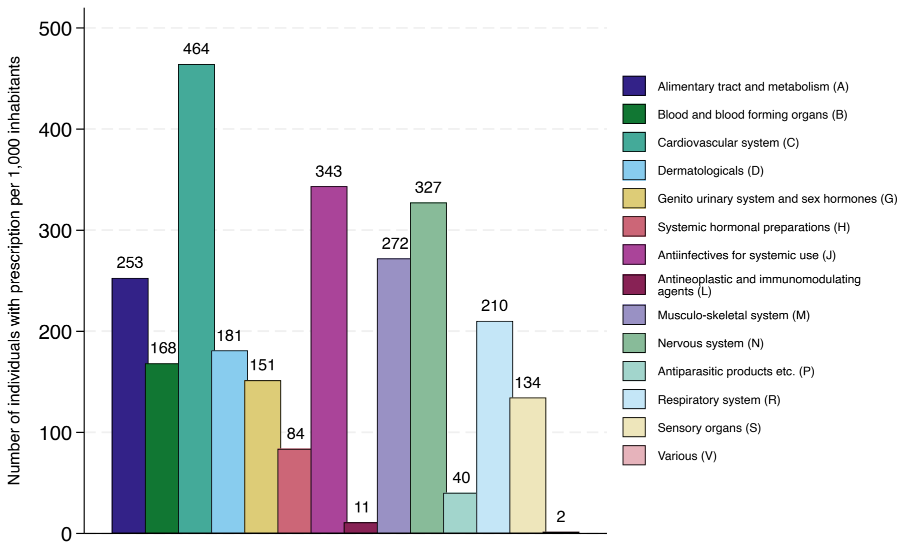
B.2 Health and Healthcare Utilization Outcome Variables

Figure B.2: Prevalence of CVD Medicine, Painkillers, and Antidepressants



The figure shows the number of individuals aged 60 who have redeemed at least one prescription of the given category in 2010 per 1,000 inhabitants. The medicine categories follow the ATC codes of www.whocc.no. Source: https://medstat.dk (The Danish Health Data Authority).

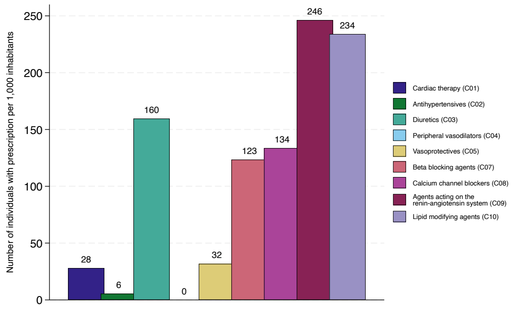
Figure B.3: Prevalence of All Medicine Categories (ATC 1st level)



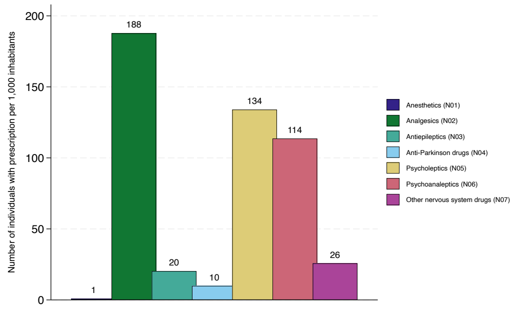
The figure shows the number of individuals aged 60 who have redeemed at least one prescription of the given category in 2010 per 1,000 inhabitants. The medicine categories follow the ATC codes of www.whocc.no. Source: https://medstat.dk (The Danish Health Data Authority).

Figure B.4: Decomposition of Selected Medicine Categories

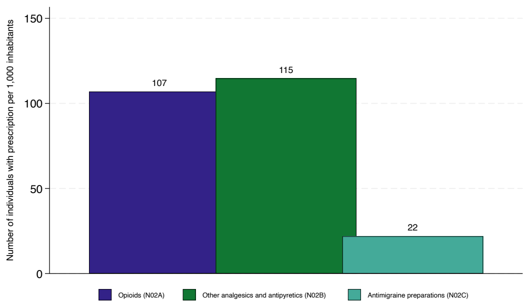
(a) Cardiovascular System



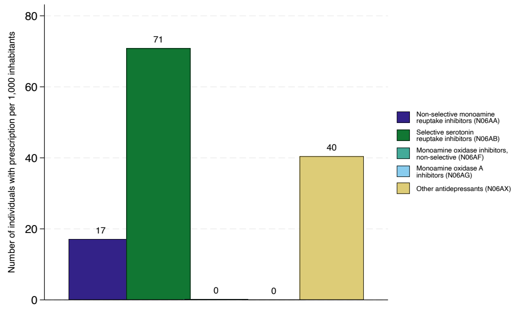
(b) Nervous System



(c) Analgesics

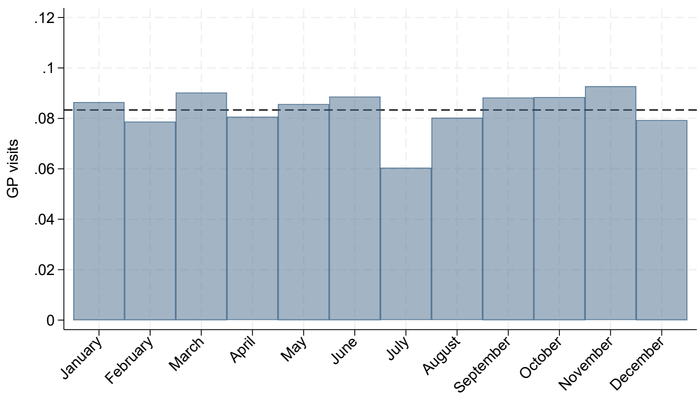


(d) Antidepressants



The figure shows the number of individuals aged 60 who have redeemed at least one prescription of the given category in 2010 per 1,000 inhabitants. Panel (a) shows the prevalence across subcategories for prescriptions for the cardiovascular system (C01-C10), panel (b) for prescriptions for the nervous system (N01-N07), panel (c) for analgesics (N02A-N02C), and panel (d) for antidepressants (N06AA-N06AX). The medicine groups are not mutually exclusive, as individuals may redeem prescriptions for multiple categories in 2010. The medicine categories follow the ATC codes of www.whocc.no.
Source: https://medstat.dk (The Danish Health Data Authority).

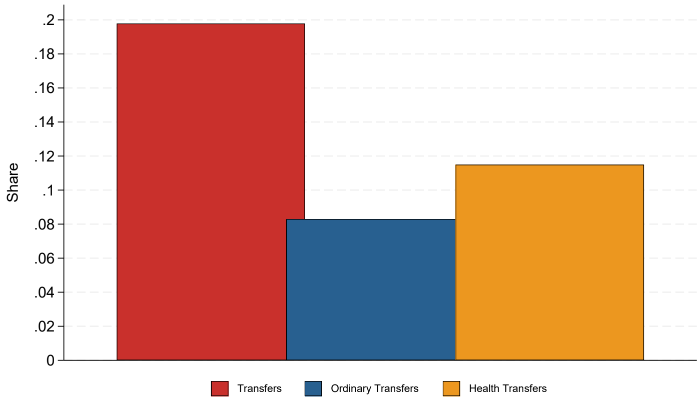
Figure B.5: GP Visits by Month



The figure shows the share of GP visits by month for the years 2005-2010. The dotted line at 0.083 represents the expected share if GP visits were evenly distributed throughout the year.

B.3 Details on Transfer Types

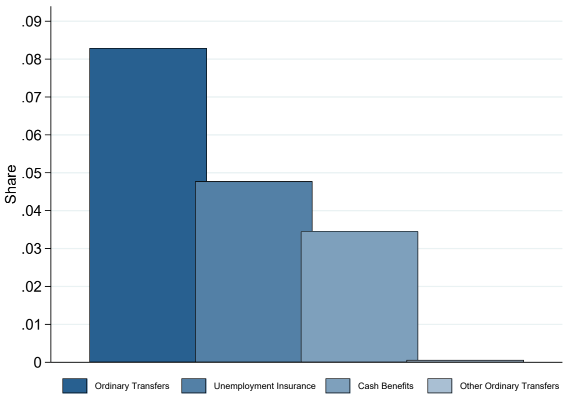
Figure B.6: Transfers Excl. Retirement Benefits



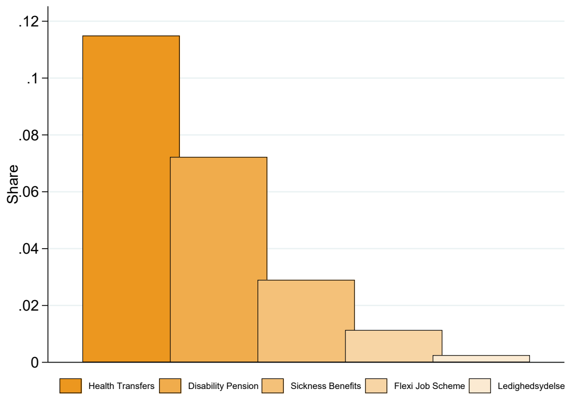
The figure shows the share of individuals aged 25-65 on different types of transfers in the period 2000-2010. The shares are calculated as the average share of time (measured in weeks) spent on a given transfer type during the year. "Transfers" corresponds to the sum of "Ordinary Transfers" and "Health Transfers". Only the most important transfer is recorded for each individual in a given week. See www.dst.dk (in Danish) for a detailed description.

Figure B.7: Ordinary Transfers and Health-Related Transfers

(a) Ordinary Transfers



(b) Health Transfers



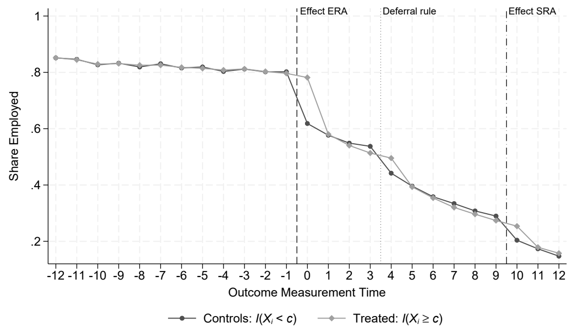
The figure shows the share of individuals aged 25-65 on different types of transfers in the period 2000-2010. The bars "Ordinary Transfers" (left) and "Health Transfers" (right) are equivalent to the bars of the same names in Figure B.6. "Ordinary Transfers" is the sum of "Unemployment Insurance", "Cash Benefits", and "Other Ordinary Transfers", while "Health Transfers" is the sum of "Disability Pension", "Sickness Benefits", "Flexi Job Scheme", and "Ledighedsydelse".

B.4 Descriptives: Labor Force Participation

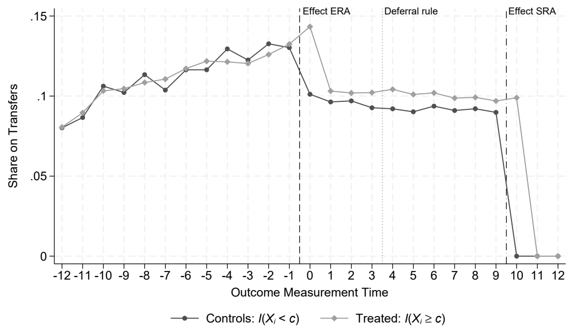
In Figure B.8, we plot the average share of individuals who are employed or receiving other transfers over time, spanning from six years before to six and a half years after the ERA increase, by whether they are born before or after the cutoff.

Figure B.8: Labor Force Participation Over Time

(a) Employment



(b) Other Transfers



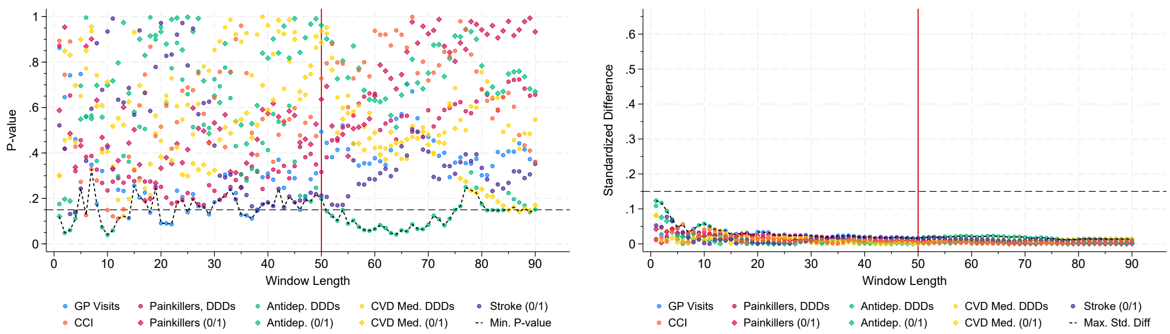
The figure shows the share employed in panel (a) and the share on other transfers in panel (b) in half-year intervals from 6 years before to 6.5 years after the ERA increase. See Section 2.3.2 for a detailed description of labor market outcomes. "Controls" are individuals born prior to July 1, while "Treated" are born on July 1 or later.

C Alternative Specifications and Robustness

C.1 Window Choice

Figure C.1: Choosing the Bandwidth, w

(a) Difference-in-means Test (b) Absolute Standardized Difference



The figure shows the P-values of difference-in-means tests (left) and absolute standardized differences (right) of nine predetermined health covariates at the cutoff at bandwidths ranging from 1 to 90 days. The smallest P-value of the difference-in-means tests and the largest absolute standardized difference are indicated with dashed lines in each figure. The horizontal dashed lines indicate a P-value of 0.15 (left), and an absolute standardized difference of 0.25 (right), being the levels recommended by Cattaneo et al. (2023) and Stuart and Rubin (2008), respectively. The vertical line indicates a bandwidth of 50 days. Variables are measured for a four-year period prior to the reform announcement (age 52½-56½ for 1954:2Q-1954:3Q cohort and age 51½-55½ for 1955:2Q-1955:3Q cohort).

C.2 The Increased ERA's Effect on Labor Force Participation - Conditional on Previous Employment

In Table C.1, we repeat the estimations in Section 4.1, but condition the sample on being employed in the half-year period prior to the ERA increase, $t = -1$ (age 60-60½ for 1954:Q2-1954:Q3 cohort and age 61-61½ for 1955:Q2-1955:Q3 cohort). We define individuals as being employed if they have more than three months of full-time equivalent work in the half-year period. We impose this sample restriction to show that the spillover effects on ordinary transfers and self-supporting are not entirely driven by individuals who were previously out of employment. As seen from Table C.1, the effect on self-supporting (Column (5)) is still sizable and statistically significant, however, we find no effect on the binary indicator of being on transfers (Column (2)). We caution, however, that data on employment (from BFL, monthly) and data on transfers (from DREAM, weekly) are from two different data sources, and that previously employed individuals may increase their reliance on transfers without this becoming their primary status during the half-year period. To address this, we use the ratios of time spent receiving ordinary transfers (Column (3)) and health-related transfers (Column (4)) instead of a binary indicator for whether these transfers are an individual's primary status. We then find that there is a positive effect on primarily ordinary transfers (49%), but also for health transfers (11%), even though there is no significant change in the probability of having other transfers as the primary status.

Table C.1: Effects on LFP - Conditional on Employment in Previous Period

	(1) Employment (0/1)	(2) Transfers (0/1)	(3) Ordinary transfers (pct.)	(4) Health transfers (pct.)	(5) Self-supporting (0/1)
$\mathbb{1}[X_i \geq c]$	0.203***	-0.000351	0.00654***	0.00539*	0.00498***
	(0.00468)	(0.00222)	(0.00147)	(0.00291)	(0.00148)
Mean	0.76	0.02	0.01	0.05	0.01
Pct. Change	26.66	-1.46	49.00	10.66	63.10
N	18,940	18,940	18,940	18,940	18,940

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on labor force participation outcomes estimated using Equation (1). All outcomes are measured the first half-year following the ERA increase, $t = 0$ (age 60½-61 for 1954:Q2-1954:Q3 cohort and age 61½-62 for 1955:Q2-1955:Q3 cohort). The outcomes in Columns (1) and (2) are measured as binary variables equal to 1 if the individual has more than three months of full-time equivalent work (Column (1)) and receives transfers (excluding retirement benefits) for more than three months and is not employed (Column (2)). The outcomes in Columns (3)-(4) are ratios of the period with ordinary transfers (Column (3)) and health transfers (Column (4)). Self-supporting in Column (5) is when an individual is neither employed nor on transfers for three or more months. The sample is conditional on being employed, as defined in Column (1), at $t = -1$. "Mean" refers to the average for individuals born before the cutoff, c (July 1). Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

C.3 The Continuity-based Framework

In Table C.2 and Figure C.2, we employ the continuity-based framework instead of the local randomization framework, following Cattaneo et al. (2019). We choose the mean squared error optimal bandwidth for each outcome separately and weight observations using the triangular kernel. We choose a local polynomial of order 1 to construct the point estimates and a local polynomial of order 2 to construct the bias correction. Table C.2 displays both the conventional and bias-corrected robust estimates. Figure C.2 illustrates the conventional estimates. Overall, this framework leads to effects somewhat quantitatively similar to the ones in Table 2, but with much larger imprecision.

Table C.2: The Continuity Framework: Effects on Health and Healthcare Utilization Outcomes

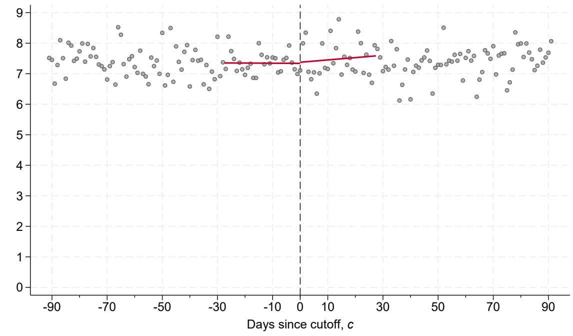
	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
Conventional	0.0329	0.865	7.027	-18.49
	(0.264)	(7.733)	(10.53)	(54.04)
Robust	0.0154	-0.217	9.788	-29.84
	(0.315)	(8.452)	(12.36)	(64.02)
Pct. Change	0.45	1.27	10.97	-2.52
Bandwidth	27.45	28.87	24.56	21.06
Efficient N	12,876	13,385	11,455	10,065
N	43,658	43,658	43,658	43,658

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

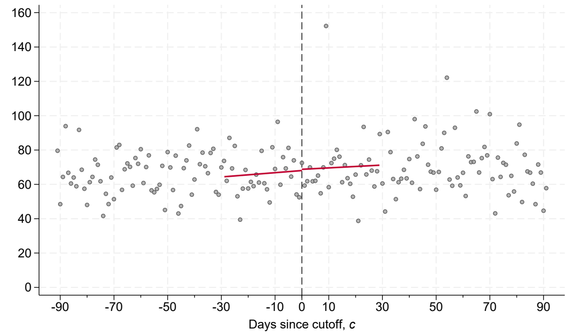
Effects on health outcomes estimated under the continuity-based framework. All outcomes denoted in the column headers are measured over a two-year period post-ERA. The row denoted “Conventional” shows conventional estimates with conventional standard errors in parentheses, while the row denoted “Robust” shows the robust bias-corrected estimates with robust bias-corrected standard errors in parentheses. The bandwidth is chosen to optimize the mean squared error and observations are weighted using the triangular kernel. The rows “Bandwidth” and “Efficient N” denote the chosen bandwidth and the resulting included number of observations, while the row “N” refers to the full number of observations included in choosing the bandwidth. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

Figure C.2: Continuity Assumption: Effects on Health Outcomes

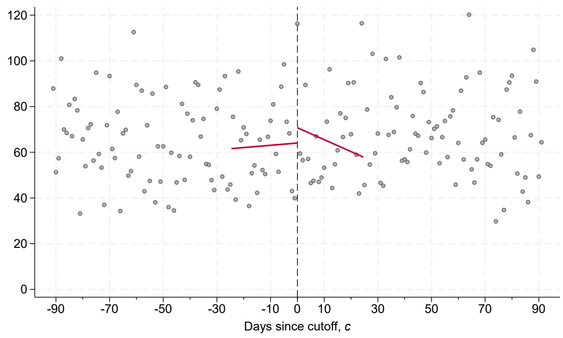
(a) GP Visits



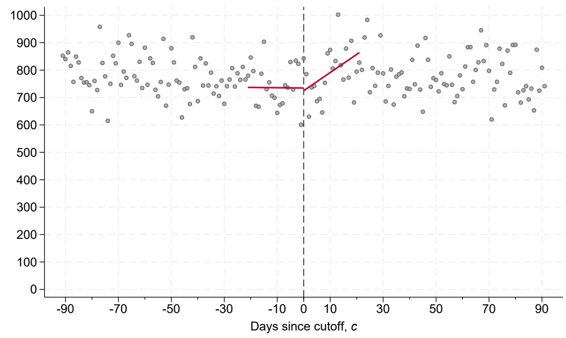
(b) Painkillers DDDs



(c) Antidepressants DDDs



(d) CVD Medicine DDDs

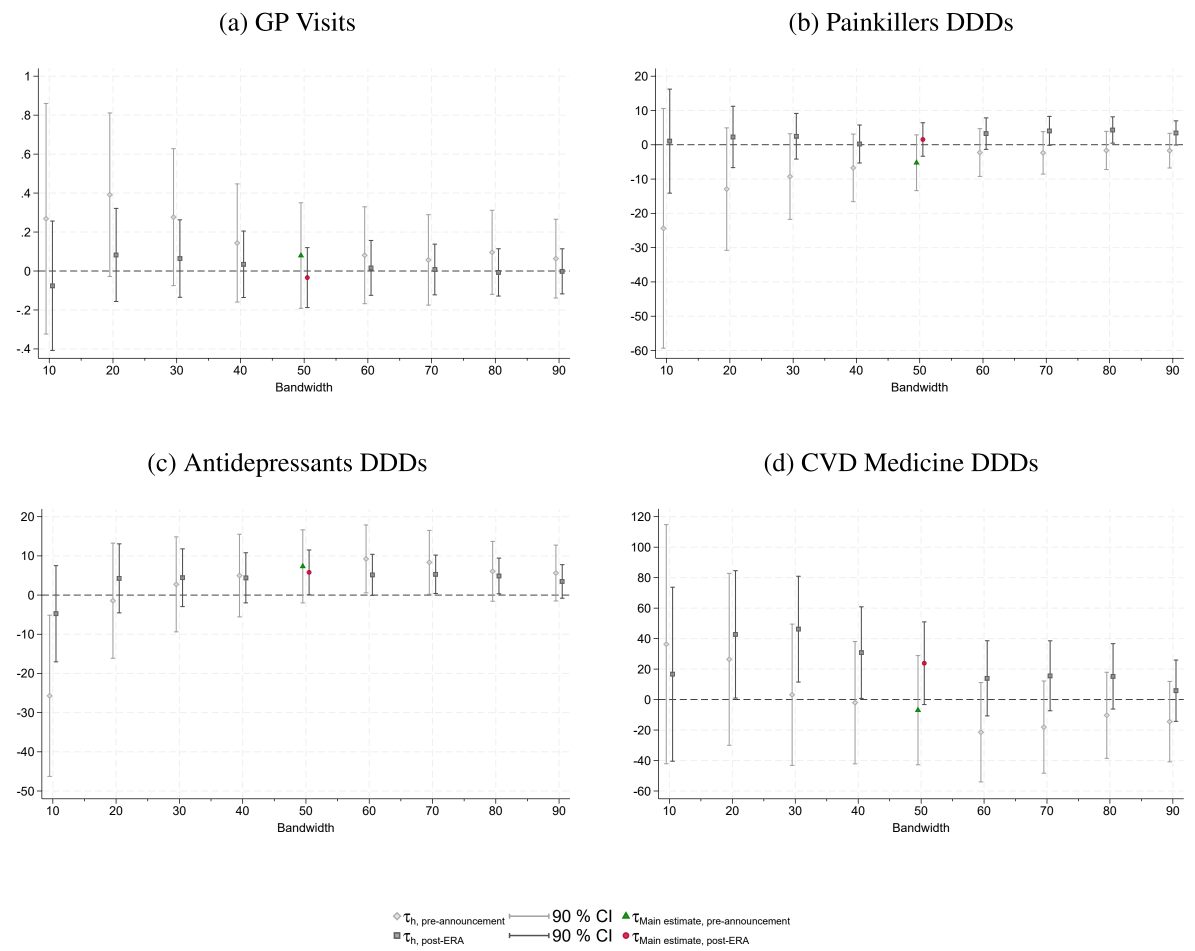


The figure illustrates the results from Table C.2, estimating the effect on health and healthcare utilization under the continuity-based framework. The grey dots show averages of the outcome variables for each day of the running variable for 90 days on each side of the cutoff (July 1). "Local linear fit" (the red line) refers to the local polynomial of order 1 fitted on the data within the mean squared error optimal bandwidth. All outcomes are measured for a two-year post-ERA period.

C.4 Robustness to Bandwidth Choice

Figure C.3 shows the main estimation results as a function of the chosen bandwidth, *w*, ranging from 10 to 90, for pre-announcement estimates and post-ERA estimates. In our main specification, we choose a bandwidth of 50, as indicated by the green triangle (pre-announcement) and red circle (post-ERA) in Figure C.3. Overall, the main estimates are relatively stable across varying bandwidths, especially for increasing bandwidths. As expected, the estimates are very imprecise for low bandwidths, especially in the range 10-30.

Figure C.3: Robustness to Bandwidth Choice



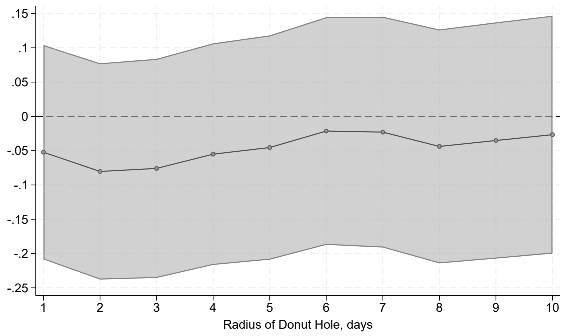
The figure illustrates the results from estimating Equation (1) for varying bandwidth, *w*. Our main specification uses a bandwidth of 50, indicated by the green triangle (pre-announcement) and red circle (post-ERA). Pre-announcement health outcomes are measured four years before reform announcement (age 52½-56½ for cohort 1954:Q2-1954:Q3 and age 51½-55½ for cohort 1955:Q2-1955:Q3). Post-ERA health outcomes are measured for a two-year period post-ERA (age 60½-62½ for cohort 1954:Q2-1954:Q3 and age 61½-63½ for cohort 1955:Q2-1955:Q3). Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms. The gray capped spikes show 90% confidence intervals calculated with robust standard errors.

C.5 Donut Hole Specifications

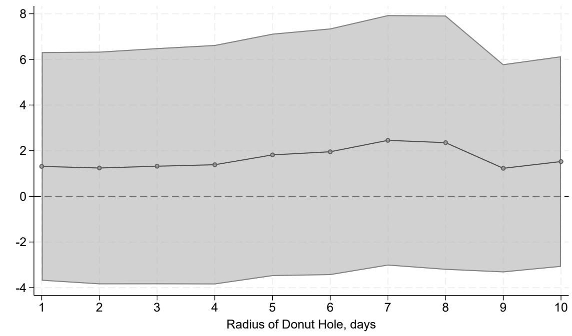
Figure C.4 shows the main estimation results where we exclude observations in a radius, h, around the cutoff, where h varies from 1 to 10. Overall, the main estimates are relatively stable across varying exclusion radiuses.

Figure C.4: Donut Hole Specifications

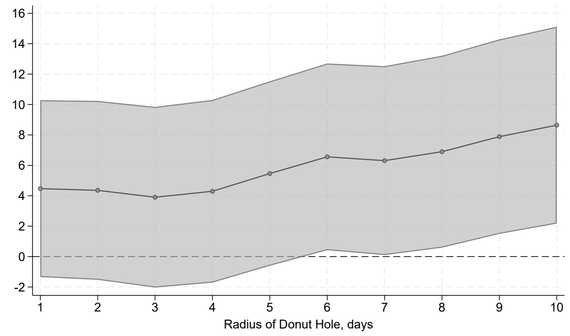
(a) GP Visits



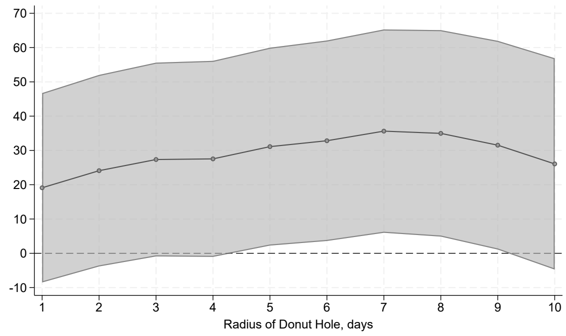
(b) Painkillers DDDs



(c) Antidepressants DDDs



(d) CVD Medicine DDDs



The figure illustrates the results from estimating Equation (1) excluding a radius, h around the cutoff, where h varies from 1 to 10. Health outcomes are measured for a two-year period post-ERA. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms. The gray areas show 90% confidence intervals calculated with robust standard errors.

C.6 Sensitivity to Outliers

In Table C.3, we test for the sensitivity of outliers by winsorizing our outcomes at the 1^{st} and 99^{th} percentile. We find similar results to Table 2, showing that our results are not driven by outliers.

Table C.3: Effects on Health and Healthcare Utilization Outcomes

	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
$\mathbb{1}[X_i \geq c]$	-0.00266	-0.617	4.714	21.51
	(0.0836)	(2.037)	(3.037)	(15.53)
Mean	7.22	63.42	58.95	743.43
Pct. Change	-0.04	-0.97	8.00	2.89
N	23,694	23,694	23,694	23,694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on health and healthcare utilization outcomes estimated using Equation (1). All outcomes denoted in the column headers are measured a two-year period after the ERA increase. All outcomes have been winsorised at the 1^{st} and 99^{th} percentile, and the outcome variables are described in detail in Section 2.3.1. "Mean" refers to the average for the group born before July 1 (i.e., prior to the cutoff). Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

C.7 RD-DD

We estimate two versions of a Regression Discontinuity Difference-in-Differences (RD-DD) using either (i) earlier cohorts, or (ii) individuals who did not contribute to the early retirement scheme at age 55 as our control group.

C.7.1 RD-DD: Earlier Cohorts

We estimate RD-DD (i) with the following equation for individual i at relative time since ERA increase r:

$$Y_{ir} = \beta_{0,r} + \beta_{1,r} \cdot \mathbb{1}[t = \tau] + \beta_{2,r} \cdot \mathbb{1}[X_i(t) \geq c(t)] + \beta_{3,r} \cdot \mathbb{1}[t = \tau] \times \mathbb{1}[X_i(t) \geq c(t)] + Z_i \delta_r + \varepsilon_{ir}, \quad \text{(C1)}$$

where cutoff, $c(t)$, and running variable, $X_i(t)$ depend on the specific cohort, t, similarly to Equation (1). Here, $\mathbb{1}[t = \tau]$ is an indicator variable equal to one if the cohort is affected by the increased ERA. In Table C.4 and Figure C.5, we include two treated (1954:2Q-1954:3Q, 1955:2Q-1955:3Q) and two control cohorts (1952:2Q-1952:3Q, 1953:2Q-1953:3Q). In this case, the cutoff, $c(t)$ is July 1, 1954 or July, 1955, for the treated cohorts and July 1, 1952 or July 1, 1953, for the control cohorts. In this way, the coefficient on $\mathbb{1}[t = \tau] \times \mathbb{1}[X_i(t) \geq c(t)]$ is the additional effect from being after July 1 in a treated cohort.

The outcome, Y_{ir}, is measured for a two-year period after the ERA (at age 60½-62½ for 1954:2Q-1954:3Q, age 61½-63½ for 1955:2Q-1955:3Q, and age 60-62 for the control cohorts). In this way, the coefficient on $\mathbb{1}[t = \tau] \times \mathbb{1}[X_i(t) \geq c(t)]$ is the additional effect from being to the right of a cutoff, $c(t)$, in a treated cohort, τ. Control variables, Z_i, are measured at age 55 if not otherwise specified and include: a male dummy, a marriage dummy, dummy for short cycle tertiary education or higher, average net wealth age 48 to 52 squared, and average disposable income age 48 to 52 squared.

The results in Tables C.4 and C.5 and Figures C.5 and C.6 fail to substantiate any general cutoff-specific impacts. Furthermore, the magnitudes and patterns in Tables C.4 and C.5 largely mirror those reported in Table 2, but with higher imprecision, due to additionally estimating general cutoff-specific effects.

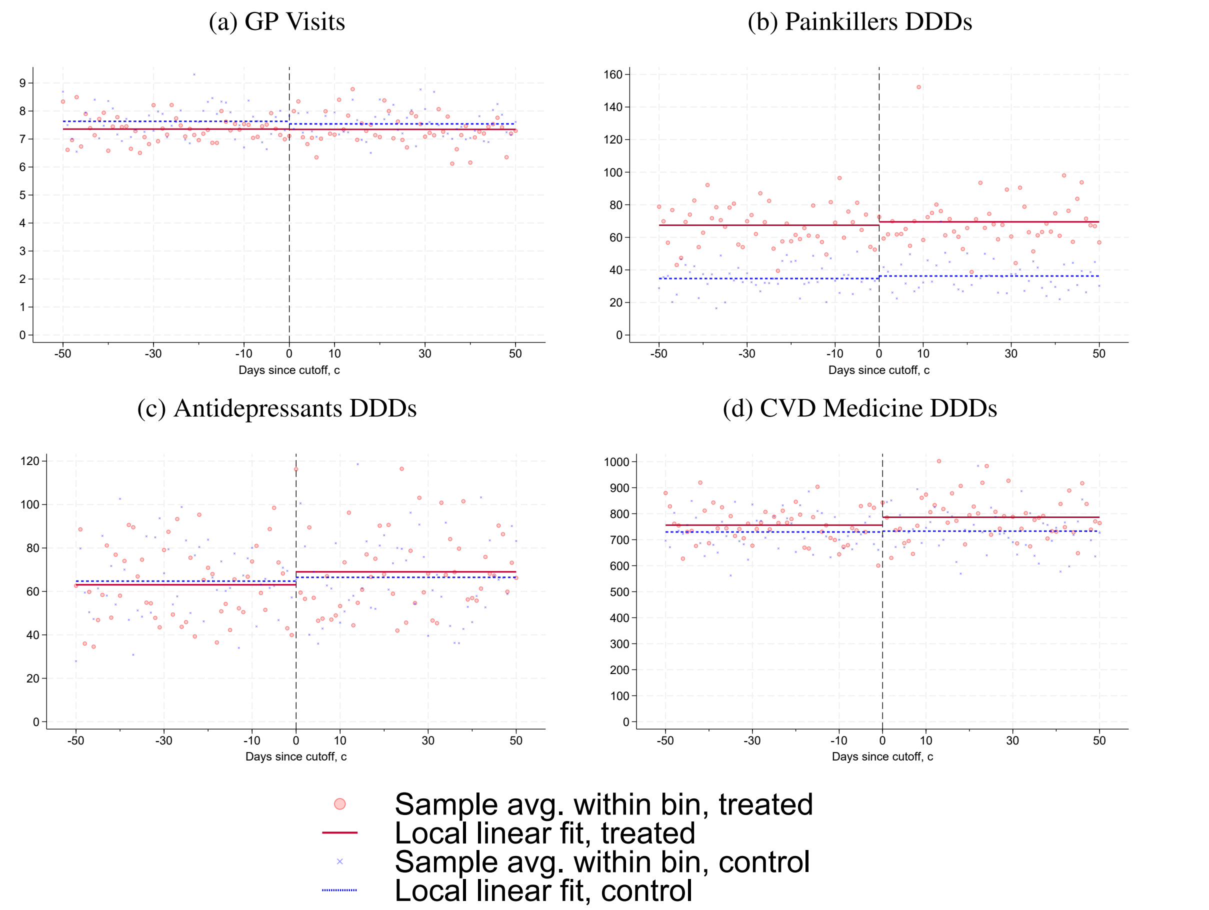
Table C.4: RD-DD: Effects on Health Outcomes

	(1) GP visits	(2) DDD painkillers	(3) DDD antidep.	(4) DDD CVD medicine
$\mathbb{1}[t=\tau]$	-0.190*	34.26***	-0.226	39.43**
	(0.0997)	(2.097)	(3.353)	(15.83)
$\mathbb{1}[X_i \geq c]$	-0.0931	1.522	1.666	3.322
	(0.103)	(1.665)	(3.423)	(15.58)
$\mathbb{1}[X_i \geq c] \times \mathbb{1}[t=\tau]$	0.0641	0.252	4.259	21.75
	(0.139)	(3.406)	(4.877)	(22.70)
Pct. Additional Change	0.87	0.37	6.75	2.88
Bandwidth	50	50	50	50
N	47,288	47,288	47,288	47,288

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

The table shows the results of estimating Equation (C1) on health and healthcare utilization outcomes. The estimation sample includes two treated and two control cohorts. See Appendix C.7.1 for details. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, and average of net wealth and disposable income (ages 48–52), including squared terms.

Figure C.5: RD-DD: Effects on Health Outcomes



The figure illustrates the results from Table C.4, estimating Equation (C1) on health and healthcare utilization outcomes. See Appendix C.7.1 for details.

C.7.2 RD-DD: Early Retirement Contribution

In this RD-DD, we use individuals who at age 55 did not contribute to the early retirement scheme as control group as they should not be expected to be affected by the ERA increase. The sample consists of individuals born in 1954:Q2-1954:Q3 and 1955:Q2-1955:Q3, corresponding to Column (2) in Table B.1.

We estimate RD-DD (ii) using for individual i at relative time since ERA increase r:

$$Y_{ir} = \beta_{0,r} + \beta_{1,r} \cdot \omega_i + \beta_{2,r} \cdot \mathbb{1}\left[X_i(t) \geq c(t)\right] + \beta_{3,r} \cdot \omega_i \times \mathbb{1}\left[X_i(t) \geq c(t)\right] + Z_i\delta_r + \varepsilon_{ir}, \tag{C2}$$

where cutoff, $c(t)$, and running variable, $X_i(t)$ depend on the specific cohort, t, and ω_i is a dummy variable equal to one if an individual has paid early retirement contributions at age 55. The cutoff, $c(t)$ is July 1, 1954 or July, 1955. The coefficient on $\omega_i \times \mathbb{1}\left[X_i(t) \geq c(t)\right]$ is the additional effect from being born on or after July 1 *and* having contributed to the early retirement scheme. For both treated and control, the outcome is measured for the two-year period following the ERA increase. Control variables, Z_i, are measured at age 55 if not otherwise specified and include: a male dummy, a marriage dummy, dummy for short cycle tertiary education or higher, average net wealth age 48 to 52 squared, average disposable income age 48 to 52 squared, and a dummy for being born in 1955 (1954 as reference group)

In Table C.5 and Figure C.6, we show the results. First, we note from Figure C.6 that individuals who did not contribute to the early retirment scheme have higher healthcare utilization both before and after the July 1-cutoff for all health-related outcomes (especially for use of painkillers). We find similar results to Table 2, except for antidepressants where we instead find a negative estimate close to zero. However, given the large standard errors, the confidence interval spans from -17 to 16 DDDs, thereby including our main estimate of 6 DDDs. For CVD medicine, we find an effect of 40 DDD (which is even larger than our main estimate of 26 DDDs).

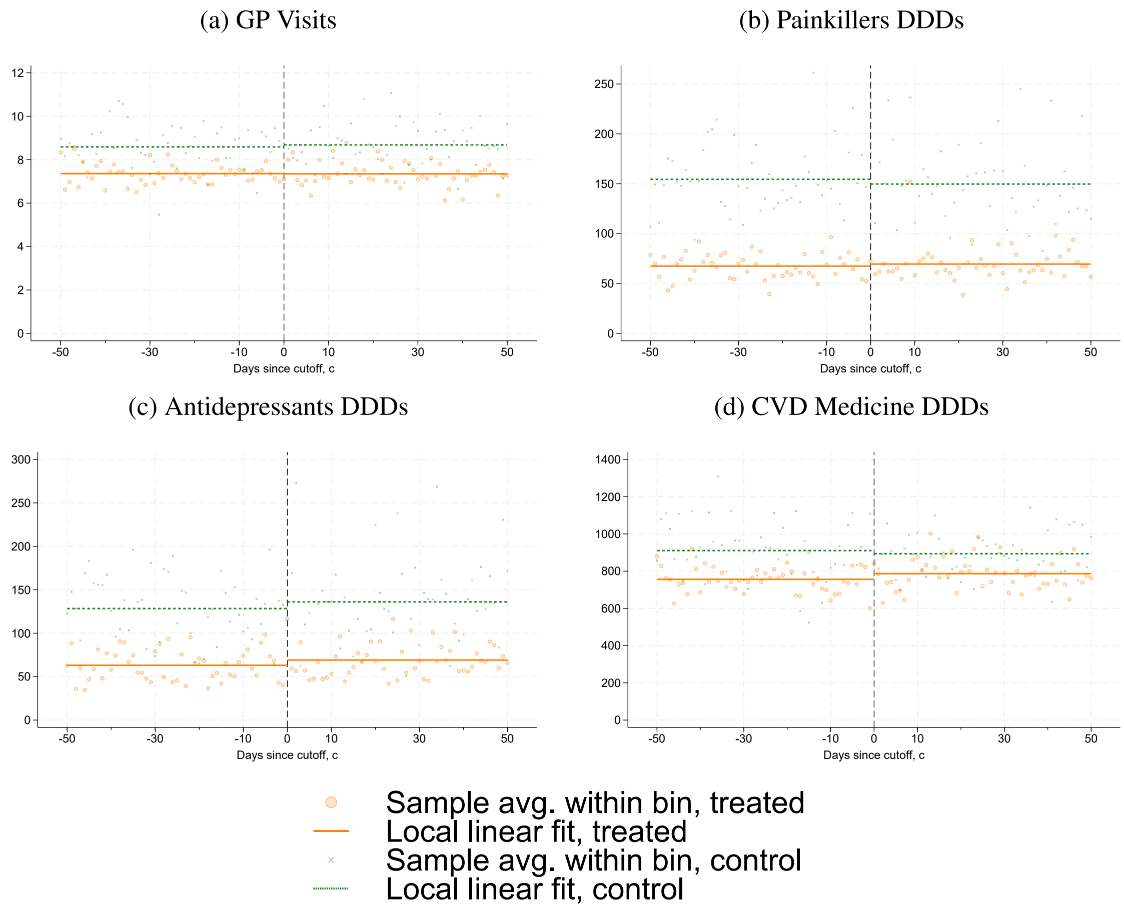
Table C.5: Pooled RD-DD: Effects on Health Outcomes: ERA payment

	(1)	(2)	(3)	(4)
	GP visits	DDD painkillers	DDD antidep.	DDD CVD medicine
ω_i	-1.039***	-81.08***	-57.61***	-114.9***
	(0.147)	(5.558)	(5.862)	(23.86)
$\mathbb{1}[X_i \geq c]$	0.0411	-6.205	6.291	-16.02
	(0.186)	(7.452)	(7.923)	(28.29)
$\mathbb{1}[X_i \geq c] \times \omega_i$	-0.0761	7.568	-0.500	40.41
	(0.208)	(8.018)	(8.652)	(32.76)
Pct. Additional Change	-1.03	11.22	-0.79	5.34
Bandwidth	50	50	50	50
N	34,263	34,263	34,263	34,263

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

The table shows the results of estimating Equation (C2) on health and healthcare utilization outcomes. The estimation sample includes individuals who have paid into the ERA scheme at age 55 (treated) and those who have not (control). See Appendix C.7.2 for details. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

Figure C.6: Pooled RD-DD: Effects on Health Outcomes: ERA payment



The figure illustrates the results from Table C.5, estimating Equation (C2) on health and healthcare utilization outcomes. The estimation sample includes individuals who have paid into the ERA scheme at age 55 (treated) and those who have not (control). See Appendix C.7.2 for details.

D Additional Results

D.1 Additional Health and Healthcare Utilization Outcomes

We consider the following additional health and healthcare utilization outcomes. First, we study the binary outcomes of the prescription drugs as described in Section 2.3.1. Then, we turn to consider the more severe health-related outcomes as described below.

Charlson Comorbidity Index (CCI): The CCI, as introduced by Charlson et al. (1987), is a well-recognized objective health measure, which has previously been used in the economics literature (see, e.g., Contoyannis et al. (2005), Nielsen (2016) or Nielsen (2019)). The CCI aggregates unique ICD-10 diagnoses, weighted for comorbidity severity, capturing conditions ranging from congestive heart failure to metastatic solid tumors. We use the same ICD-10 diagnoses and weights as in Christensen et al. (2011) and Nielsen (2019). Note that the data include only diagnoses from hospital contacts, excluding those solely from GP or psychiatric consultations unless also recorded upon a hospital contact. We interpret CCI as a compound measure of health, similar to a health stock in the line of Grossman (1972). The largest drawback of this measure is that it may take a long time before effects materialize, and that detection of diagnoses may differ between groups of individuals.

Stroke: We consider stroke (binary) as this can be seen as being very close to an objective health measure. We create an indicator variable for having a stroke equal to 1 if an individual has had hospital contact with the ICD-10 code being either subarachnoid haemorrhage (I60), intracerebral haemorrhage (I61), cerebral infarction (I63), or stroke, not specified as haemorrhage or infarction (I64). As having a stroke is life-threatening and requires emergency care, individuals are advised to call for an ambulance as soon as symptoms appear. Without medical care, strokes can be deadly. Each year, approximately 12,000 Danes experience a stroke (The Danish Clinical Quality Program, 2024). Among the risk factors are high stress levels and an unhealthy lifestyle, e.g., poor diet and low physical activity.

Mortality: We consider mortality as this is a direct health measure which is not affected by choice. We create a binary indicator for mortality by using whether an individual is present in the population register (BEF), however, this is only measured annually. Therefore, we cannot measure mortality as precisely as the other health-related outcomes, and we are hence only studying mortality over a two-, four-, and six-year period following the ERA increase.

We show the effect of ERA increase on the additional health and healthcare utilization outcomes over different time periods in Table D.1 and dynamic effects by half-yearly intervals on Figure D.1. In Table D.2, we study the effect of a half-year increase in the ERA on mortality, where we find no effects.

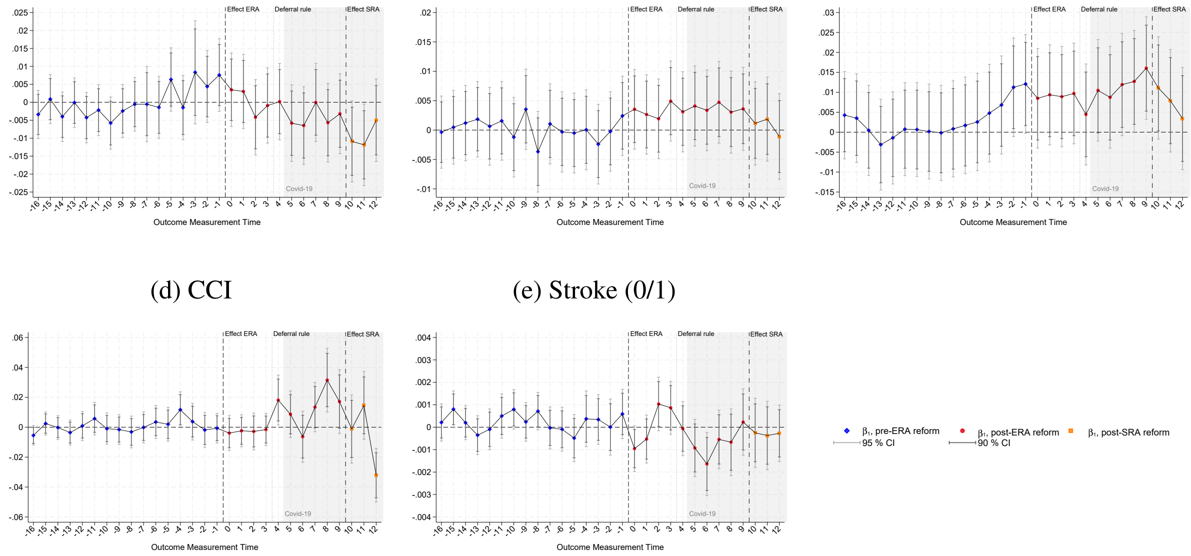
Table D.1: Effects on Health and Healthcare Utilization Outcomes

	(1) Painkillers Usage (0/1)	(2) Antidep. Usage (0/1)	(3) CVD Med. Usage (0/1)	(4) Stroke (0/1)	(5) CCI
Panel A: Pre-announcement (4 years)					
$\mathbb{1}[X_i \geq c]$	0.00174	0.0000843	-0.00192	0.00137	0.00513
	(0.00542)	(0.00432)	(0.00631)	(0.00113)	(0.0162)
Mean	0.24	0.13	0.42	0.01	0.27
Pct. Change	0.73	0.06	-0.46	19.21	1.89
N	24,286	24,286	24,286	24,286	24,286
Panel B: Post-announcement					
$\mathbb{1}[X_i \geq c]$	0.00819	-0.00428	0.0133**	0.00193	0.00986
	(0.00620)	(0.00453)	(0.00638)	(0.00146)	(0.0289)
Mean	0.40	0.15	0.52	0.01	0.58
Pct. Change	2.02	-2.85	2.59	16.08	1.71
N	24,237	24,237	24,237	24,237	24,237
Panel C: Post-ERA (half-year)					
$\mathbb{1}[X_i \geq c]$	0.00347	0.00354	0.00848	-0.000956*	-0.00393
	(0.00522)	(0.00345)	(0.00636)	(0.000519)	(0.00595)
Mean	0.20	0.08	0.41	0.00	0.07
Pct. Change	1.70	4.71	2.06	-46.49	-5.88
N	23,694	23,694	23,694	23,694	23,694
Panel D: Post-ERA (2 years)					
$\mathbb{1}[X_i \geq c]$	0.00254	0.00401	0.00864	0.000626	-0.0107
	(0.00623)	(0.00386)	(0.00645)	(0.00107)	(0.0147)
Mean	0.37	0.10	0.49	0.01	0.26
Pct. Change	0.68	4.14	1.75	9.76	-4.12
N	23,694	23,694	23,694	23,694	23,694
Panel E: Post-ERA (4 years)					
$\mathbb{1}[X_i \geq c]$	-0.00262	0.00632	0.00857	-0.00116	0.0226
	(0.00644)	(0.00421)	(0.00638)	(0.00156)	(0.0337)
Mean	0.50	0.12	0.57	0.02	0.75
Pct. Change	-0.53	5.39	1.51	-7.66	3.01
N	23,694	23,694	23,694	23,694	23,694
Panel F: Post-ERA (6 years)					
$\mathbb{1}[X_i \geq c]$	-0.0103	0.00246	0.00992	-0.00151	0.0828
	(0.00635)	(0.00451)	(0.00619)	(0.00193)	(0.0605)
Mean	0.59	0.14	0.64	0.02	1.56
Pct. Change	-1.75	1.76	1.56	-6.48	5.30
N	23,694	23,694	23,694	23,694	23,694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses

Effects on additional health and healthcare utilization outcomes estimated using Equation (1). The outcome variables are described in detail in Appendix D.1. "Mean" refers to the average for the group born before July 1 (i.e., prior to the cutoff). Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

Figure D.1: Dynamic Effects on Additional Health and Healthcare Utilization Outcomes



(a) Painkillers Usage (0/1) (b) Antidep. Usage (0/1) (c) CVD Med. Usage (0/1)

(d) CCI (e) Stroke (0/1)

The figure illustrates the results from estimating Equation (1) on additional health and healthcare utilization outcomes in half-year intervals from 6 years before the ERA increase to 6.5 years after. The outcome variables are described in detail in Appendix D.1. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms. The gray capped spikes show 90% and 95% confidence intervals calculated with robust standard errors.

Table D.2: Effects on Mortality

	(1)	(2)	(3)
	2 years	4 years	6 years
$\mathbb{1}[X_i \geq c]$	0.000348	-0.000111	-0.000280
	(0.00145)	(0.00207)	(0.00266)
Pct. Change	2.82	-0.43	-0.64
Mean	0.01	0.03	0.04
N	23,694	23,694	23,694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

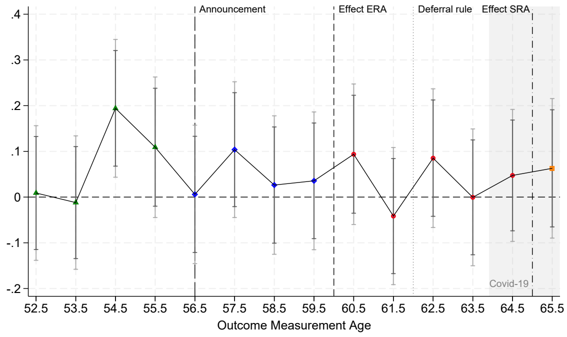
Effects on mortality estimated using Equation (1). Outcomes are measured for a two-year period after the ERA increase in Column (1), four-year period in Column (2), and six-year period in Column (3). "Mean" refers to the average for the group born before July 1 (i.e., prior to the cutoff). Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, birth year 1955 (vs. 1954), and average of net wealth and disposable income (ages 48–52), including squared terms.

D.2 Cutoff-Specific Results

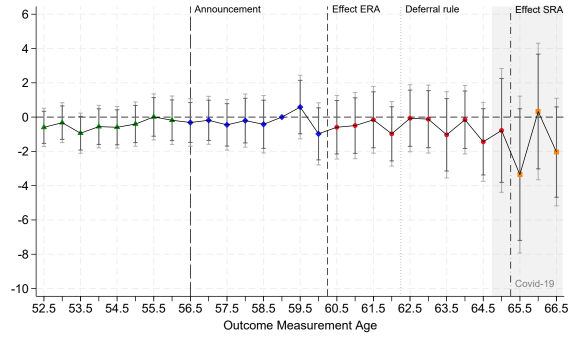
We study the effects separately for each cutoff in Appendix D.2 to examine whether the impact of the half-year ERA increase is consistent across groups. We start by showing the cutoff-specific dynamic effects over time for the two summer cutoffs used in our pooled analysis in Figure D.2 (1954:Q2–1954:Q3) and Figure D.3 (1955:Q2–1955:Q3). Second, we show the cutoff-specific effects of a half-year ERA increase for pre-announcement, post-announcement, and post-ERA periods for 1954:Q2–1954:Q3 (Table D.3), 1955:Q2–1955:Q3 (Table D.4), winter cutoffs, 1953:Q4-1954:Q1 (Table D.5) and 1954:Q4-1955:Q1 (Table D.6), and the later summer cutoff, 1956:Q2–1956:Q3 (Table D.7). While the bandwidth of 50 days was chosen for a pooled sample for 1954:Q2–1954:Q3 and 1955:Q2–1955:Q3, we apply the same 50-day bandwidth in all cutoff-specific analyses. Reassuringly, we do not find any health effects in the pre-announcement period, supporting the validity of this choice. The one exception is the 1954:Q2–1954:Q3 cutoff, where we observe a positive effect on antidepressants prior to the announcement but no effects post-announcement.

Figure D.2: Effects on Health and Healthcare Utilization Outcomes Over Time: 1954:Q2-1954:Q3

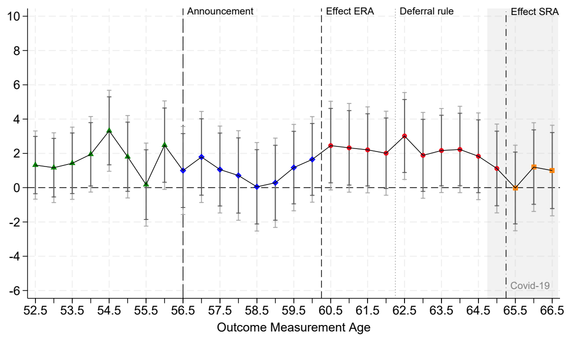
(a) GP Visits



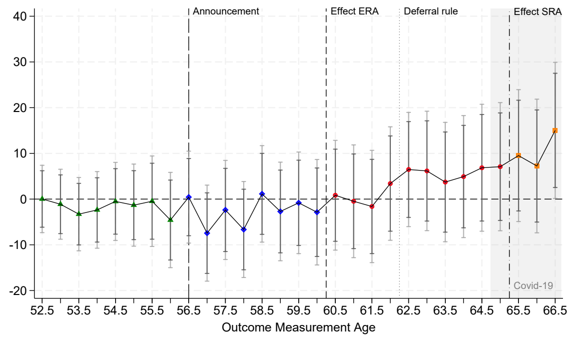
(b) Painkillers DDDs

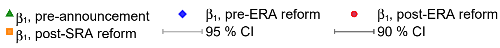


(c) Antidep. DDDs



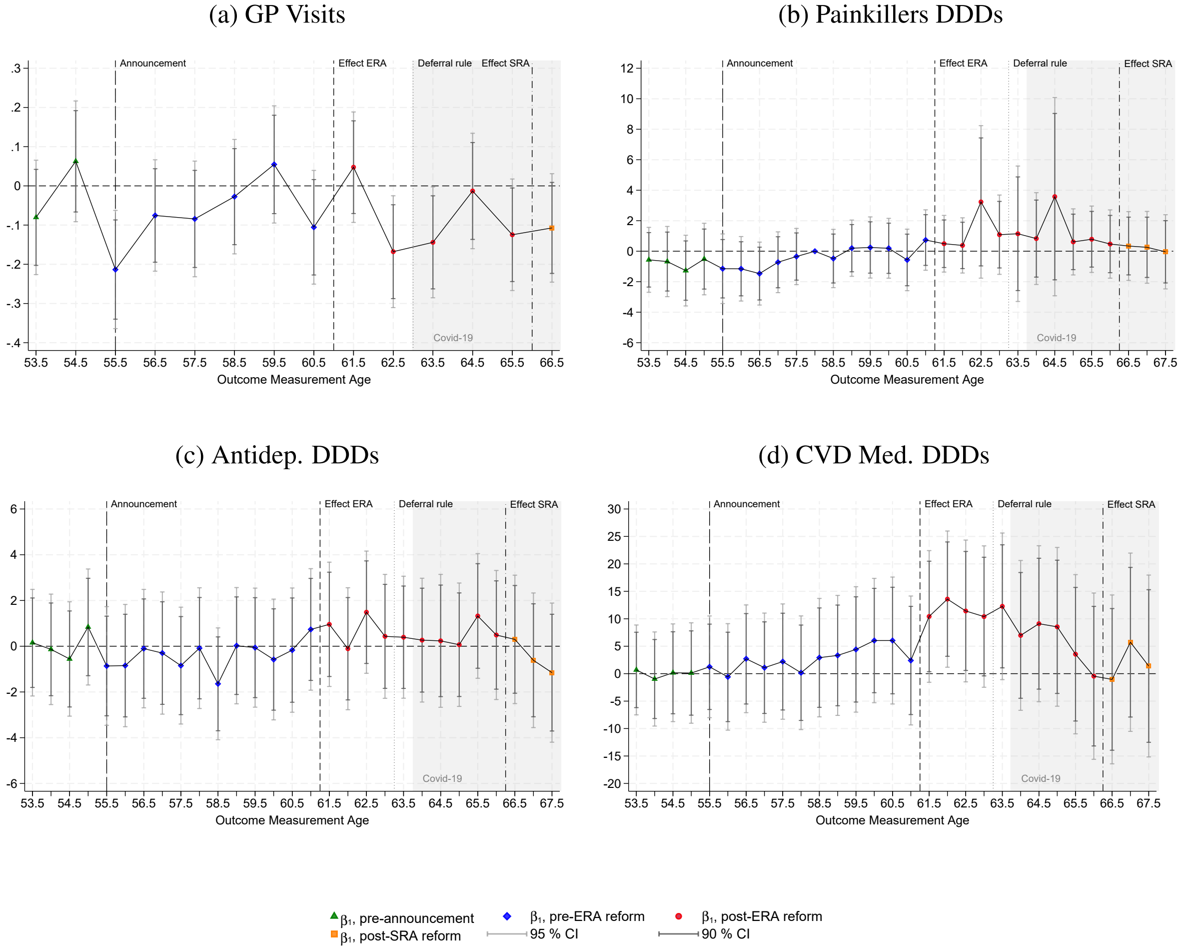
(d) CVD Med. DDDs





The figure illustrates the results from estimating Equation (1) on main health and healthcare utilization outcomes in half-year intervals (yearly intervals for GP visits) from 8 years before the ERA increase to 6.5 years after for cohort 1954:Q2-1954:Q3 with cutoff July 1, 1954. Outcome variables are described in detail in Section 2.3.1. The gray bar indicates that the estimates include years affected by the Covid-19 pandemic, which may influence healthcare utilization. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, and average of net wealth and disposable income (ages 48–52), including squared terms. The gray capped spikes show 90% and 95% confidence intervals calculated with robust standard errors.

Figure D.3: Effects on Health and Healthcare Utilization Outcomes Over Time: 1955:Q2-1955:Q3



The figure illustrates the results from estimating Equation (1) on main health and healthcare utilization outcomes in half-year intervals (yearly intervals for GP visits) from 8 years before the ERA increase to 6.5 years after for cohort 1955:Q2-1955:Q3 with cutoff July 1, 1955. Outcome variables are described in detail in Section 2.3.1. The gray bar indicates that the estimates include years affected by the Covid-19 pandemic, which may influence healthcare utilization. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, and average of net wealth and disposable income (ages 48–52), including squared terms. The gray capped spikes show 90% and 95% confidence intervals calculated with robust standard errors.

Table D.3: Effects on Health and Healthcare Utilization Outcomes: 1954:Q2-1954:Q3

	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
Panel A: Pre-announcement (4 years)				
$\mathbb{1}[X_i \geq c]$	0.300	-3.609	13.59*	-13.87
	(0.240)	(4.627)	(7.939)	(31.92)
Mean	13.95	49.21	101.71	830.34
Pct. Change	2.15	-7.33	13.36	-1.67
N	12,135	12,135	12,135	12,135
Panel B: Post-announcement				
$\mathbb{1}[X_i \geq c]$	0.167	-2.022	7.594	-21.80
	(0.243)	(4.963)	(9.239)	(39.24)
Mean	14.08	69.01	126.42	1217.40
Pct. Change	1.19	-2.93	6.01	-1.79
N	12,086	12,086	12,086	12,086
Panel C: Post-ERA (short-run)				
$\mathbb{1}[X_i \geq c]$	0.0934	-0.592	2.453*	0.843
	(0.0785)	(0.947)	(1.319)	(6.126)
Mean	3.61	16.21	14.96	179.79
Pct. Change	2.59	-3.65	16.40	0.47
N	11,878	11,878	11,878	11,878
Panel D: Post-ERA (2 years)				
$\mathbb{1}[X_i \geq c]$	0.0470	-2.256	8.906*	1.682
	(0.138)	(3.584)	(4.768)	(23.11)
Mean	7.32	67.07	58.39	752.92
Pct. Change	0.64	-3.36	15.25	0.22
N	11,878	11,878	11,878	11,878
Panel D: Post-ERA (4 years)				
$\mathbb{1}[X_i \geq c]$	0.123	-3.624	18.02**	21.27
	(0.247)	(7.115)	(9.164)	(46.11)
Mean	14.80	139.97	114.91	1592.79
Pct. Change	0.83	-2.59	15.69	1.34
N	11,878	11,878	11,878	11,878
Panel E: Post-ERA (6 years)				
$\mathbb{1}[X_i \geq c]$	0.233	-8.690	22.12*	49.28
	(0.336)	(11.70)	(13.30)	(69.75)
Mean	21.56	227.28	174.63	2580.39
Pct. Change	1.08	-3.82	12.66	1.91
N	11,878	11,878	11,878	11,878

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on health and healthcare utilization outcomes estimated using Equation (1) for cohort 1954:Q2-1954:Q3 with cutoff July 1, 1954. The outcome variables are described in detail in Section 2.3.1. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, and average of net wealth and disposable income (ages 48–52), including squared terms.

Table D.4: Effects on Health and Healthcare Utilization Outcomes: 1955:Q2-1955:Q3

	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
Panel A: Pre-announcement (4 years)				
$\mathbb{1}[X_i \geq c]$	-0.147	-6.866	1.296	-0.0461
	(0.226)	(8.752)	(8.084)	(29.66)
Mean	13.91	48.70	110.04	724.18
Pct. Change	-1.05	-14.10	1.18	-0.01
N	12,151	12,151	12,151	12,151
Panel B: Post-announcement				
$\mathbb{1}[X_i \geq c]$	-0.381	-3.886	-4.062	39.47
	(0.336)	(9.032)	(13.42)	(56.99)
Mean	21.44	132.10	205.58	1754.58
Pct. Change	-1.78	-2.94	-1.98	2.25
N	12,151	12,151	12,151	12,151
Panel C: Post-ERA (short-run)				
$\mathbb{1}[X_i \geq c]$	0.0477	0.491	0.957	10.43*
	(0.0719)	(0.948)	(1.386)	(6.120)
Mean	3.62	16.58	17.17	177.86
Pct. Change	1.32	2.96	5.57	5.86
N	11,816	11,816	11,816	11,816
Panel D: Post-ERA (2 years)				
$\mathbb{1}[X_i \geq c]$	-0.116	5.195	2.792	45.71*
	(0.126)	(4.734)	(5.053)	(23.55)
Mean	7.40	67.86	67.91	759.30
Pct. Change	-1.57	7.66	4.11	6.02
N	11,816	11,816	11,816	11,816
Panel D: Post-ERA (4 years)				
$\mathbb{1}[X_i \geq c]$	-0.262	11.40	3.869	83.38*
	(0.222)	(11.61)	(9.753)	(47.86)
Mean	14.23	147.90	136.26	1681.54
Pct. Change	-1.84	7.70	2.84	4.96
N	11,816	11,816	11,816	11,816
Panel E: Post-ERA (6 years)				
$\mathbb{1}[X_i \geq c]$	-0.479	13.24	5.313	92.38
	(0.311)	(14.21)	(14.35)	(73.20)
Mean	20.67	238.44	208.07	2786.90
Pct. Change	-2.32	5.55	2.55	3.31
N	11,816	11,816	11,816	11,816

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on health and healthcare utilization outcomes estimated using Equation (1) for cohort 1955:Q2-1955:Q3 with cutoff July 1, 1955. The outcome variables are described in detail in Section 2.3.1. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, and average of net wealth and disposable income (ages 48–52), including squared terms.

Table D.5: Effects on Health and Healthcare Utilization Outcomes: 1953:Q4-1954:Q1

	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
Panel A: Pre-announcement (4 years)				
$\mathbb{1}[X_i \geq c]$	0.114	1.019	-6.671	34.63
	(0.236)	(4.322)	(8.433)	(33.26)
Mean	14.23	45.32	122.03	855.29
Pct. Change	0.80	2.25	-5.47	4.05
N	11,869	11,869	11,869	11,869
Panel B: Post-announcement				
$\mathbb{1}[X_i \geq c]$	-0.161	1.280	-2.352	24.25
	(0.191)	(3.470)	(7.210)	(30.32)
Mean	11.15	37.14	102.52	910.79
Pct. Change	-1.44	3.45	-2.29	2.66
N	11,782	11,782	11,782	11,782
Panel C: Post-ERA (short-run)				
$\mathbb{1}[X_i \geq c]$	-0.0925	0.889	0.149	4.645
	(0.0778)	(0.889)	(1.243)	(6.053)
Mean	3.73	13.04	15.53	172.16
Pct. Change	-2.48	6.82	0.96	2.70
N	11,617	11,617	11,617	11,617
Panel D: Post-ERA (2 years)				
$\mathbb{1}[X_i \geq c]$	-0.114	2.597	-0.452	5.606
	(0.138)	(3.391)	(4.658)	(22.93)
Mean	7.49	57.15	62.04	730.63
Pct. Change	-1.53	4.54	-0.73	0.77
N	11,617	11,617	11,617	11,617
Panel D: Post-ERA (4 years)				
$\mathbb{1}[X_i \geq c]$	-0.0675	4.605	-0.501	34.52
	(0.243)	(6.705)	(8.948)	(45.71)
Mean	15.04	121.38	122.64	1534.44
Pct. Change	-0.45	3.79	-0.41	2.25
N	11,617	11,617	11,617	11,617
Panel E: Post-ERA (6 years)				
$\mathbb{1}[X_i \geq c]$	-0.132	9.996	-5.073	57.46
	(0.329)	(10.57)	(13.20)	(69.21)
Mean	22.27	192.80	186.49	2477.75
Pct. Change	-0.59	5.18	-2.72	2.32
N	11,617	11,617	11,617	11,617

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on health and healthcare utilization outcomes estimated using Equation (1) for cohort 1953:Q4-1954:Q1 with cutoff January 1, 1954. The outcome variables are described in detail in Section 2.3.1. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, and average of net wealth and disposable income (ages 48–52), including squared terms.

Table D.6: Effects on Health and Healthcare Utilization Outcomes: 1954:Q4-1955:Q1

	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
Panel A: Pre-announcement (4 years)				
$\mathbb{1}[X_i \geq c]$	-0.0329	0.329	-4.510	33.91
	(0.254)	(5.218)	(8.682)	(31.30)
Mean	14.09	42.34	115.45	754.76
Pct. Change	-0.23	0.78	-3.91	4.49
N	11,580	11,580	11,580	11,580
Panel B: Post-announcement				
$\mathbb{1}[X_i \geq c]$	-0.116	4.616	-2.274	48.78
	(0.309)	(7.135)	(11.73)	(49.70)
Mean	17.92	91.42	164.33	1468.78
Pct. Change	-0.65	5.05	-1.38	3.32
N	11,535	11,535	11,535	11,535
Panel C: Post-ERA (short-run)				
$\mathbb{1}[X_i \geq c]$	-0.201**	0.557	0.430	0.625
	(0.0839)	(0.981)	(1.312)	(6.228)
Mean	3.88	16.33	15.70	185.79
Pct. Change	-5.20	3.41	2.74	0.34
N	11,302	11,302	11,302	11,302
Panel D: Post-ERA (2 years)				
$\mathbb{1}[X_i \geq c]$	-0.238	1.668	2.252	2.320
	(0.147)	(3.567)	(4.907)	(23.59)
Mean	7.68	66.48	62.16	765.11
Pct. Change	-3.10	2.51	3.62	0.30
N	11,302	11,302	11,302	11,302
Panel D: Post-ERA (4 years)				
$\mathbb{1}[X_i \geq c]$	-0.384	3.350	4.611	9.347
	(0.249)	(6.888)	(9.516)	(47.92)
Mean	14.89	137.66	125.14	1669.02
Pct. Change	-2.58	2.43	3.68	0.56
N	11,302	11,302	11,302	11,302
Panel E: Post-ERA (6 years)				
$\mathbb{1}[X_i \geq c]$	-0.347	8.226	4.922	1.640
	(0.333)	(10.50)	(14.05)	(73.87)
Mean	21.40	221.88	191.10	2760.29
Pct. Change	-1.62	3.71	2.58	0.06
N	11,302	11,302	11,302	11,302

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on health and healthcare utilization outcomes estimated using Equation (1) for cohort 1954:Q4-1955:Q1 with cutoff January 1, 1955. The outcome variables are described in detail in Section 2.3.1. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, and average of net wealth and disposable income (ages 48–52), including squared terms.

Table D.7: Effects on Health and Healthcare Utilization Outcomes: 1956:Q2-1956:Q3

	(1) GP visits	(2) Painkillers DDDs	(3) Antidep. DDDs	(4) CVD Med. DDDs
Panel A: Pre-announcement (4 years)				
$\mathbb{1}[X_i \geq c]$	0.219	-0.699	8.678	30.93
	(0.238)	(3.412)	(8.102)	(28.37)
Mean	13.24	37.32	98.71	643.42
Pct. Change	1.65	-1.87	8.79	4.81
N	11,803	11,803	11,803	11,803
Panel B: Post-announcement				
$\mathbb{1}[X_i \geq c]$	-0.120	-10.34	11.66	74.44
	(0.424)	(10.19)	(17.88)	(75.07)
Mean	27.62	193.30	251.15	2331.31
Pct. Change	-0.43	-5.35	4.64	3.19
N	11,803	11,803	11,803	11,803
Panel C: Post-ERA (short-run)				
$\mathbb{1}[X_i \geq c]$	-0.109	-0.141	0.615	15.81**
	(0.0713)	(1.136)	(1.355)	(6.421)
Mean	3.37	18.26	16.34	199.78
Pct. Change	-3.23	-0.77	3.76	7.91
N	11,512	11,512	11,512	11,512
Panel D: Post-ERA (2 years)				
$\mathbb{1}[X_i \geq c]$	-0.109	-0.243	1.209	43.29*
	(0.128)	(4.320)	(5.102)	(25.38)
Mean	6.76	78.99	66.25	882.81
Pct. Change	-1.61	-0.31	1.83	4.90
N	11,512	11,512	11,512	11,512
Panel D: Post-ERA (4 years)				
$\mathbb{1}[X_i \geq c]$	-0.0694	4.539	4.775	82.69
	(0.225)	(10.97)	(9.937)	(51.51)
Mean	13.16	173.54	132.43	1944.64
Pct. Change	-0.53	2.62	3.61	4.25
N	11,512	11,512	11,512	11,512

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

Effects on health and healthcare utilization outcomes estimated using Equation (1) for cohort 1956:Q2-1956:Q3 with cutoff July 1, 1956. The outcome variables are described in detail in Section 2.3.1. Control variables (measured at age 55 unless noted) include dummies for gender, being married, higher education, and average of net wealth and disposable income (ages 48–52), including squared terms.

D.3 Characterization of Compliers: Increase in Statutory Retirement Age (SRA)

We repeat the complier analysis, this time focusing on SRA compliers, i.e., individuals who worked in response to the increase in the statutory retirement age . We compare them to always-takers, who are employed regardless, and never-takers, who remain non-employed irrespective of the reform.

E Heterogeneous Treatment Effects Estimated by Causal Forest

To examine heterogeneous treatment effects of the ERA extension on our four main outcomes, we apply the causal forest algorithm following Athey and Imbens (2016), Wager and Athey (2018), Athey et al. (2019), implemented via the `grf` R-package. The purpose of the causal forest approach is to predict heterogeneity in the causal treatment effects and estimate Conditional Average Treatment Effects (CATEs), defined as $\mathbb{E}[Y_{1i} - Y_{0i} | X_i = x]$, where Y is the outcome of interest, and X is a vector of observable characteristics.

We include 41 characteristics in X for the causal forest: gender (dummy), marital status (dummy), education (categorical variable with levels five levels: less than upper secondary, upper secondary, short-cycle teritary, bachelor or equivalent, and master or equivalent education), average disposable income at age 48-52, average net wealth at age 48-52, occupational level (categorical variable with four levels: managerial, high-skilled, medium-skilled, or other work), experience, employment (dummy), public employee (dummy), white collar worker (dummy), GP visits (measured over four years before reform announcement), painkillers (DDDs, measured over four years before reform announcement), antidepressants (DDDs, measured over four years before reform announcement), CVD medicine (DDDs, measured over four years before reform announcement), 10 industry dummies (following the standard groupings of Statistics Denmark) and 17 occupational characteristics from $O^{*}NET$[34] and BIBB[35]: routine, codifiability, time pressure, competition, importance being exact, automation, freedom, impact, hazardous, cramped, angry people, conflict, coordination, contact with others, physical, decision-making freedom, time management. All occupational variables are measured at age 55, meaning that unemployed individuals will have missing values. We replace missing values with 0. This is consistent with the algorithm's use of data partitioning to maximize treatment effect heterogeneity, which allows splits at zero.[36]

We use the default parameter values except for the number of trees and the minimum node size. The minimum node size sets a requirement for each leaf to contain a minimum number of observations. We tune the minimum node size by fitting 100 "mini forests" (500 trees each) and sampling 1,000 parameter values. The optimal minimum node sizes range from 350 to 650 across outcomes. Larger node sizes improve precision and reduce computation time relative to the default (5).

The final causal forest prediction is a weighted average over the predictions of 100,000 trees. We choose 100,000 trees to obtain mean excess errors that are negligible compared to the mean variance

[34] Occupational Information Network, developed under sponsorship of U.S. Department of Labor

[35] Bundesinstitut für Berufsbildung, i.e., the German Federal Institute for Vocational Education and Training.

[36] The GRF command can also handle missing variables implicitly by using the missing incorporated in attributes criterion.

estimates for all of the four main outcomes.

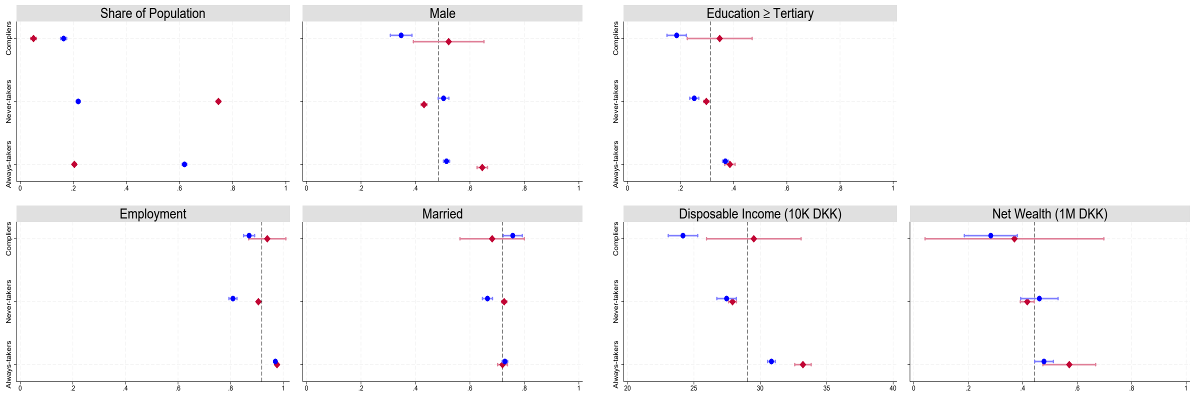
We illustrate the distributions of the estimated CATEs in Figure E.1. For all outcomes but GP visits, the estimated CATEs are greater than zero for by far the majority of individuals. This suggests that the extended ERA is unlikely to improve health for subsets of individuals through reduced healthcare utilization.

Tables E.1 to E.4 compare the average characteristics of individuals with above and below median estimated treatment effects for the four outcomes. These tables illustrate which characteristics drive the sample partitioning used to calculate the CATEs in the causal forest. Even when correcting for multiple hypothesis testing (either following List et al. (2019) in Column (4) or Holm (1979) in Column (5)), most differences are statistically significant, likely due to the large sample size. More importantly, many of the standardized differences are sizable in magnitude. For instance, for antidepressants in Table E.3, there is a 0.84 standard deviation higher share of individuals in more than median routine work in the high-CATE group compared to the low-CATE group.[37] Occupational variables appear to be key drivers of heterogeneity, especially for antidepressants, consistent with the GATEs in Figure 10.

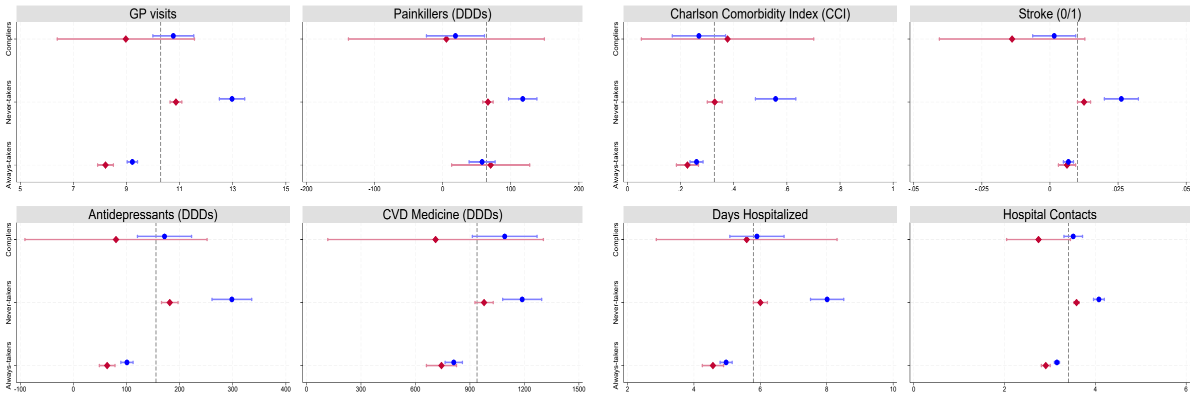
[37] Note, that for the occupational characteristics, we only include non-missing values, meaning that the difference in routine work between the high- and low-CATE groups is not driven by missing values.

Figure D.4: Characterization of ERA Compliers

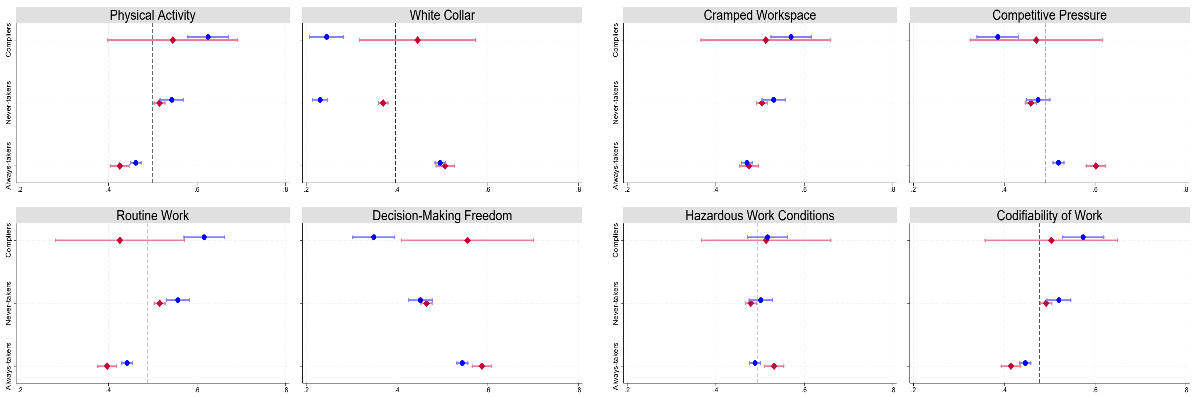
(a) Sample Share, Gender, Empl., and Marital Status (b) Educational Level, Income, and Net Wealth



(c) GP Visits and Prescription Drugs (DDDs) (d) CCI, Stroke, and Hospital Days and Contacts



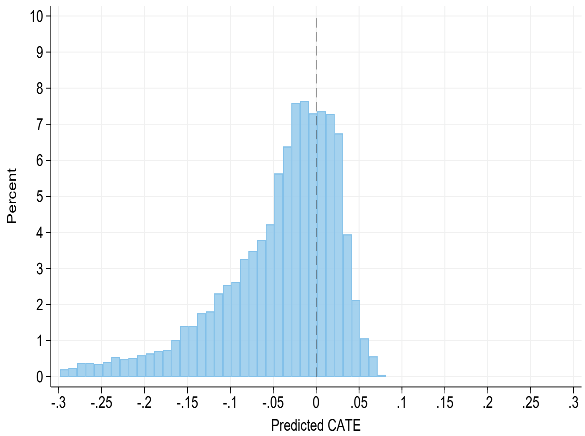
(e) Occupation: Physical, White Collar, Routine, and Freedom (f) Occupation: Cramped, Competitive, Hazardous, and Codifiability



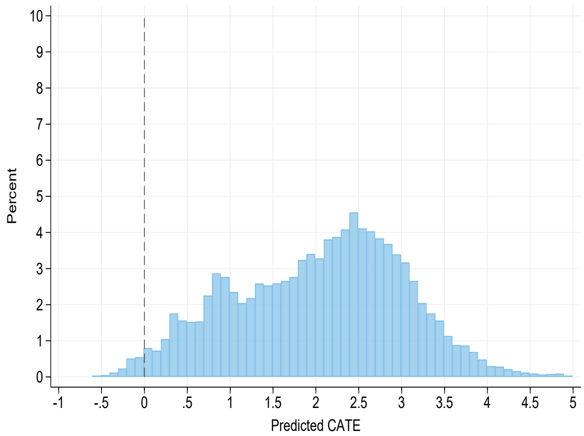
The figure characterizes the compliance strata of compliers, always-takers, and never-takers as described in Section 5.1 for the SRA increase. Complier characterization for the ERA increase (equivalent to Figure 8) is included as a reference point. The dashed, vertical lines indicate the full sample mean of each predetermined covariate. Panels (e)-(f) only include individuals employed at age 55 for which occupational characteristics can be measured. 95% confidence intervals are based on standard errors calculated asymptotically.

Figure E.1: Distribution of Estimated Conditional Average Treatment Effects (CATEs)

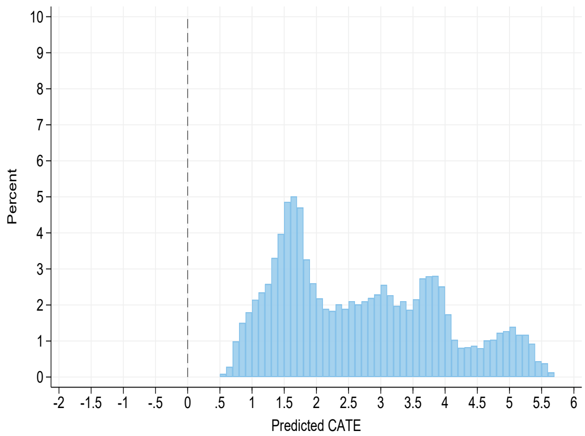
(a) GP Visits



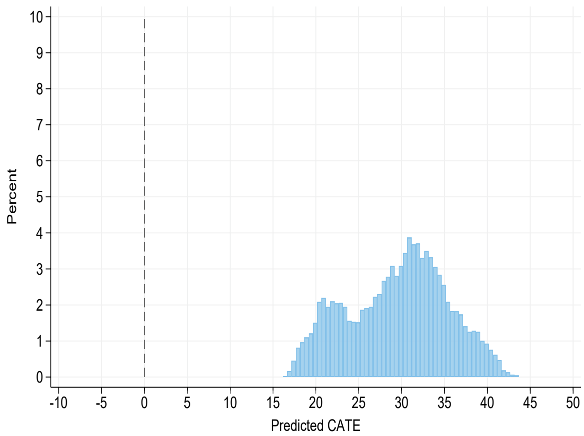
(b) Painkillers, DDDs



(c) Antidepressants, DDDs



(d) CVD Medicine, DDDs



The figure illustrates the Conditional Average Treatment Effects (CATEs) using causal forest for the four main outcomes: GP visits (Figure E.1a), painkillers (Figure E.1b), antidepressants (Figure E.1c), and CVD medicine (Figure E.1d). Each bin is conditional on containing more than five observations. The vertical dashed line in each subfigure provides a reference point at a predicted CATE of 0.

Table E.1: GP Visits: Characteristics for Individuals with High vs. Low Predicted Treatment Effects

	(1)	(2)	(3)	(4)	(5)
	Predicted Treatment Effects		Std. Diff.	MHT P-value	Holm P-value
	Below Median	Above Median	(2)-(1)	(2)-(1)	(2)-(1)
Male	0.438	0.531	0.131	0.001	0.014
Married	0.712	0.726	0.023	0.06	0.064
Education $\geq$ Tertiary	0.298	0.328	0.047	0.001	0.012
Disposable Income (10K DKK)	27.677	30.351	0.137	0.001	0.027
Net Wealth (1M DKK)	0.315	0.571	0.145	0.001	0.02
Experience, years	26.623	30.587	0.389	0.001	0.023
Employed	0.855	0.983	0.339	0.001	0.017
Pre-determined Healthcare Utilization					
GP Visits	19.994	7.825	-0.756	0.001	0.009
Antidepressants (DDDs)	163.119	52.872	-0.178	0.001	0.03
Painkillers (DDDs)	82.724	8.106	-0.136	0.001	0.029
CVD Medicine (DDDs)	1,019.873	507.854	-0.220	0.001	0.021
Occupational Characteristics†					
Managerial Work	0.065	0.031	-0.113	0.001	0.026
High-skilled Work	0.139	0.213	0.139	0.001	0.008
Medium-skilled Work	0.125	0.220	0.179	0.001	0.028
Other Work	0.671	0.536	-0.198	0.001	0.025
Public Employee	0.398	0.408	0.014	0.293	0.339
Routine	0.431	0.526	0.136	0.001	0.032
Codifiability	0.392	0.537	0.208	0.001	0.022
Time Pressure	0.499	0.498	-0.001	0.931	0.931
Competition	0.480	0.499	0.027	0.021	0.025
Importance Being Exact	0.471	0.516	0.063	0.001	0.011
Automation	0.427	0.536	0.156	0.001	0.016
Freedom	0.546	0.467	-0.112	0.001	0.007
Impact	0.560	0.457	-0.146	0.001	0.031
Hazardous	0.492	0.497	0.008	0.682	0.868
Cramped	0.483	0.504	0.031	0.018	0.024
Angry People	0.560	0.456	-0.148	0.001	0.024
Conflict	0.538	0.471	-0.094	0.001	0.015
Coordination	0.504	0.482	-0.031	0.001	0.018
Contact w. Others	0.515	0.483	-0.045	0.001	0.013
Physical	0.530	0.478	-0.074	0.001	0.01
Decision-Making Freedom	0.507	0.449	-0.083	0.001	0.033
Time Management	0.536	0.470	-0.094	0.001	0.019
Observations	11,847	11,847			

This table compares individual and occupational characteristics for individuals with above- and below-median estimated Conditional Average Treatment Effects (CATEs) on GP visits measured two years after the increased ERA. Column (3) reports standardized differences between the high- and low-CATE groups. For example, a value of 0.131 in Column (3) for "Male" indicates that the high-CATE group has a 0.13 standard deviation higher share of males than the low-CATE group. Columns (4) and (5) report p-values adjusted for multiple hypothesis testing following List et al. (2019) and Holm (1979), respectively. †: All occupational characteristics are coded as dummies. For $O^{*}NET$ and BIBB variables, dummies equal one if above the sample median. These are defined conditional on available occupational information (e.g., excluding the unemployed). Observations = 18,610.

Table E.2: Painkillers: Characteristics for Individuals with High vs. Low Predicted Treatment Effects

	(1)	(2)	(3)	(4)	(5)
	Predicted Treatment Effects		Std. Diff.	MHT P-value	Holm P-value
	Below Median	Above Median	(2)-(1)	(2)-(1)	(2)-(1)
Male	0.477	0.492	0.020	0.136	0.148
Married	0.694	0.744	0.078	0.001	0.013
Education $\geq$ Tertiary	0.342	0.284	-0.090	0.001	0.015
Disposable Income (10K DKK)	27.396	30.632	0.167	0.001	0.018
Net Wealth (1M DKK)	0.461	0.424	-0.021	0.126	0.135
Experience, years	26.505	30.705	0.414	0.001	0.01
Employed	0.838	1.000	0.439	0.001	0.016
Pre-determined Healthcare Utilization					
GP Visits	14.037	13.782	-0.014	0.319	0.366
Antidepressants (DDDs)	126.465	89.526	-0.059	0.001	0.014
Painkillers (DDDs)	40.919	49.910	0.016	0.231	0.246
CVD Medicine (DDDs)	570.518	957.209	0.165	0.001	0.012
Occupational Characteristics†					
Managerial Work	0.014	0.082	0.230	0.001	0.031
High-skilled Work	0.186	0.166	-0.039	0.001	0.033
Medium-skilled Work	0.136	0.209	0.137	0.001	0.006
Other Work	0.664	0.543	-0.176	0.001	0.008
Public Employee	0.402	0.404	0.002	0.81	0.81
Routine	0.404	0.535	0.188	0.001	0.027
Codifiability	0.356	0.548	0.278	0.001	0.029
Time Pressure	0.375	0.569	0.281	0.001	0.021
Competition	0.350	0.572	0.323	0.001	0.007
Importance Being Exact	0.409	0.548	0.198	0.001	0.019
Automation	0.334	0.581	0.362	0.001	0.03
Freedom	0.484	0.508	0.034	0.001	0.023
Impact	0.421	0.543	0.174	0.001	0.009
Hazardous	0.526	0.477	-0.069	0.001	0.024
Cramped	0.534	0.473	-0.085	0.001	0.032
Angry People	0.372	0.570	0.287	0.001	0.025
Conflict	0.421	0.542	0.172	0.001	0.011
Coordination	0.525	0.471	-0.076	0.001	0.017
Contact w. Others	0.439	0.528	0.127	0.001	0.026
Physical	0.612	0.435	-0.255	0.001	0.022
Decision-Making Freedom	0.457	0.481	0.034	0.001	0.028
Time Management	0.443	0.528	0.120	0.001	0.02
Observations	11,847	11,847			

This table compares individual and occupational characteristics for individuals with above- and below-median estimated Conditional Average Treatment Effects (CATEs) on painkillers measured two years after the increased ERA. Column (3) reports standardized differences between the high- and low-CATE groups. For example, a value of 0.02 in Column (3) for "Male" indicates that the high-CATE group has a 0.02 standard deviation higher share of males than the low-CATE group. Columns (4) and (5) report p-values adjusted for multiple hypothesis testing following List et al. (2019) and Holm (1979), respectively. †: All occupational characteristics are coded as dummies. For $O^{*}NET$ and BIBB variables, dummies equal one if above the sample median. These are defined conditional on available occupational information (e.g., excluding the unemployed). Observations = 18,610.

Table E.3: Antidepressants: Characteristics for Individuals with High vs. Low Predicted Treatment Effects

	(1)	(2)	(3)	(4)	(5)
	Predicted Treatment Effects		Std. Diff.	MHT P-value	Holm P-value
	Below Median	Above Median	(2)-(1)	(2)-(1)	(2)-(1)
Male	0.562	0.407	-0.223	0.001	0.009
Married	0.739	0.699	-0.063	0.001	0.018
Education $\geq$ Tertiary	0.427	0.199	-0.359	0.001	0.028
Disposable Income (10K DKK)	32.949	25.079	-0.420	0.001	0.016
Net Wealth (1M DKK)	0.839	0.046	-0.471	0.001	0.015
Experience, years	28.871	28.339	-0.050	0.001	0.01
Employed	0.912	0.926	0.035	0.001	0.019
Pre-determined Healthcare Utilization					
GP Visits	12.018	15.801	0.209	0.001	0.013
Antidepressants (DDDs)	90.898	125.093	0.055	0.001	0.004
Painkillers (DDDs)	33.422	57.408	0.044	0.001	0.027
CVD Medicine (DDDs)	641.105	886.622	0.105	0.001	0.003
Occupational Characteristics†					
Managerial Work	0.070	0.026	-0.147	0.001	0.025
High-skilled Work	0.270	0.081	-0.362	0.001	0.011
Medium-skilled Work	0.208	0.137	-0.133	0.001	0.017
Other Work	0.452	0.755	0.462	0.001	0.006
Public Employee	0.395	0.411	0.023	0.01	0.01
Routine	0.215	0.726	0.842	0.001	0.024
Codifiability	0.242	0.685	0.700	0.001	0.029
Time Pressure	0.584	0.424	-0.229	0.001	0.005
Competition	0.622	0.377	-0.357	0.001	0.02
Importance Being Exact	0.534	0.466	-0.097	0.001	0.032
Automation	0.480	0.501	0.029	0.007	0.008
Freedom	0.763	0.269	-0.804	0.001	0.026
Impact	0.662	0.357	-0.452	0.001	0.03
Hazardous	0.483	0.505	0.032	0.001	0.022
Cramped	0.417	0.564	0.211	0.001	0.033
Angry People	0.432	0.556	0.177	0.001	0.023
Conflict	0.549	0.454	-0.136	0.001	0.012
Coordination	0.645	0.357	-0.425	0.001	0.007
Contact w. Others	0.541	0.456	-0.122	0.001	0.021
Physical	0.401	0.586	0.266	0.001	0.014
Decision-Making Freedom	0.709	0.266	-0.700	0.001	0.008
Time Management	0.676	0.341	-0.502	0.001	0.031
Observations	11,847	11,847			

This table compares individual and occupational characteristics for individuals with above- and below-median estimated Conditional Average Treatment Effects (CATEs) on antidepressants measured two years after the increased ERA. Column (3) reports standardized differences between the high- and low-CATE groups. For example, a value of -0.223 in Column (3) for "Male" indicates that the high-CATE group has a 0.22 standard deviation lower share of males than the low-CATE group. Columns (4) and (5) report p-values adjusted for multiple hypothesis testing following List et al. (2019) and Holm (1979), respectively. †: All occupational characteristics are coded as dummies. For $O^{*}NET$ and BIBB variables, dummies equal one if above the sample median. These are defined conditional on available occupational information (e.g., excluding the unemployed). Observations = 18,610.

Table E.4: CVD Medicine: Characteristics for Individuals with High vs. Low Predicted Treatment Effects

	(1)	(2)	(3)	(4)	(5)
	Predicted Treatment Effects		Std. Diff.	MHT P-value	Holm P-value
	Below Median	Above Median	(2)-(1)	(2)-(1)	(2)-(1)
Male	0.316	0.653	0.507	0.001	0.024
Married	0.709	0.730	0.032	0.001	0.021
Education $\geq$ Tertiary	0.381	0.245	-0.209	0.001	0.023
Disposable Income (10K DKK)	29.224	28.804	-0.021	0.074	0.078
Net Wealth (1M DKK)	0.614	0.271	-0.196	0.001	0.022
Experience, years	26.883	30.327	0.335	0.001	0.015
Employed	0.838	1.000	0.440	0.001	0.007
Pre-determined Healthcare Utilization					
GP Visits	15.467	12.352	-0.172	0.001	0.009
Antidepressants (DDDs)	135.948	80.043	-0.090	0.001	0.029
Painkillers (DDDs)	54.660	36.169	-0.034	0.019	0.024
CVD Medicine (DDDs)	746.965	780.762	0.014	0.238	0.254
Occupational Characteristics†					
Managerial Work	0.034	0.061	0.089	0.001	0.016
High-skilled Work	0.214	0.138	-0.142	0.001	0.017
Medium-skilled Work	0.156	0.189	0.062	0.001	0.008
Other Work	0.596	0.612	0.023	0.037	0.04
Public Employee	0.446	0.360	-0.125	0.001	0.019
Routine	0.389	0.544	0.222	0.001	0.026
Codifiability	0.297	0.582	0.424	0.001	0.012
Time Pressure	0.340	0.589	0.366	0.001	0.014
Competition	0.343	0.576	0.341	0.001	0.027
Importance Being Exact	0.428	0.537	0.156	0.001	0.025
Automation	0.439	0.521	0.118	0.001	0.032
Freedom	0.485	0.507	0.031	0.021	0.025
Impact	0.496	0.501	0.006	0.541	0.541
Hazardous	0.273	0.622	0.529	0.001	0.018
Cramped	0.278	0.619	0.516	0.001	0.011
Angry People	0.534	0.478	-0.079	0.001	0.033
Conflict	0.586	0.448	-0.197	0.001	0.03
Coordination	0.578	0.441	-0.196	0.001	0.02
Contact w. Others	0.620	0.425	-0.282	0.001	0.031
Physical	0.357	0.581	0.326	0.001	0.028
Decision-Making Freedom	0.440	0.491	0.071	0.001	0.01
Time Management	0.556	0.464	-0.131	0.001	0.013
Observations	11,847	11,847			

This table compares individual and occupational characteristics for individuals with above- and below-median estimated Conditional Average Treatment Effects (CATEs) on CVD medicine measured two years after the increased ERA. Column (3) reports standardized differences between the high- and low-CATE groups. For example, a value of 0.507 in Column (3) for "Male" indicates that the high-CATE group has a 0.51 standard deviation higher share of males than the low-CATE group. Columns (4) and (5) report p-values adjusted for multiple hypothesis testing following List et al. (2019) and Holm (1979), respectively. †: All occupational characteristics are coded as dummies. For $O^{*}NET$ and BIBB variables, dummies equal one if above the sample median. These are defined conditional on available occupational information (e.g., excluding the unemployed). Observations = 18,610.

F Marginal Treatment Effects

In the following, we closely follow Andresen (2018) to set up the MTE in a generalized framework. Subsequently, we describe how we apply the generalized framework in our specific case.

Consider the potential outcomes, Y_0 and Y_1, in the following generalized Roy model:

$$Y_j = \mu_j(\mathcal{X}) + U_j, \qquad \text{for } j = 0,1 \tag{F1}$$

$$Y = DY_1 + (1-D)Y_0, \tag{F2}$$

$$D = \mathbb{1}\left[\mu_D(Z,\mathcal{X}) > V\right], \tag{F3}$$

where $\mathcal{X}$ are observables and Z is the instrument(s). Equation (F3) is the selection equation, interpretable as a latent index, to which the unobservable V is a negative shock. Assuming V has a continuous distribution, Equation (F3) can be written as $P(Z,\mathcal{X}) > U_D$, with $P(Z,\mathcal{X})$ being the propensity score and U_D the quantiles of V.

Identification requires the following two assumptions:

a. Conditional independence: $(U_0, U_1, V) \perp Z|\mathcal{X}$.
b. Separability: $\mathbb{E}(U_j|V,\mathcal{X},Z) = \mathbb{E}(U_j|V)$.

A model as described as Equations (F1) to (F3) with *conditional independence* is no more restrictive than the LATE model used in standard IV analysis and, in principle, it is possible to estimate MTEs without further assumptions (Vytlacil, 2002, Imbens and Angrist, 1994). However, this would require full support of the propensity score for all values of $\mathcal{X}$, for both treated and untreated samples. *Separability* allows identification of the MTE over the common support of the propensity score, unconditional on $\mathcal{X}$, and imposes that MTEs are additively separable in U_D and $\mathcal{X}$. *Separability* is a strong assumption, but commonly imposed in most applied literature estimating MTEs (e.g., Carneiro et al. (2011), Brinch et al. (2017), Cornelissen et al. (2018), Agan et al. (2023)).

Using this model, it can be shown that (Björklund and Moffitt, 1987, Heckman and Vytlacil, 2005, 2007):

$$\mathbb{E}(Y|\mathcal{X} = x, P(Z,\mathcal{X}) = p) = x\beta_0 + x(\beta_0 - \beta_1)p + K(p), \tag{F4}$$

where $K(p) = p\mathbb{E}(U_1 - U_0|U_D \leq p)$. Consequently, the MTE is given by the derivative of Equation (F4) w.r.t. p evaluated at u:

$$MTE(x,u) = (\beta_1 - \beta_0)x + k(u), \tag{F5}$$

where $k(u) = \frac{\partial K(p)}{\partial p}|_{p=u}$. The first term in Equation (F5) denotes heterogeneity in *observables* and the second term denotes heterogeneity in *unobservables*. In this way, the MTE can be interpreted as the mean return to treatment for individuals at a particular margin of indifference.

Thus, we estimate MTEs by first estimating Equation (F3) to obtain the propensity score with the following equation:

$$p(D_i = 1) = \mu(Z_i, \mathcal{X}_i), \tag{F6}$$

where D_i is an indicator variable equal to 1 if individual i is employed at the half-year of the ERA extension, $\mathcal{X}_i$ is a set of covariates, Z_i is the set of excluded instruments, and $\mu(\cdot)$ is the logit link function. $\mathcal{X}_i$ includes average wealth and disposable incomes at ages 48-52, including squared terms, and dummies for gender, being married, and education level, as well as healthcare utilization measured four years before announcement of the reform, and occupational characteristics measured at age 55. Z_i includes the binary variable for being born after July 1, 1954, $\mathbb{1}[X_i \geq c]$, as well as $\mathbb{1}[X_i \geq c]$ interacted with the wealth and disposable income variables in $\mathcal{X}_i$ and their squared terms.

Then we specify the unknown function $K(p)$ in Equation (F4) as a polynomial of degree $L+1$:

$$K(p) = \sum_{l=1}^{L} \pi_l \frac{p\left(p^l - 1\right)}{l+1}, \tag{F7}$$

where we choose $L = 3$ in our main specification. This corresponds to MTE curves being third-order polynomials.

Estimating MTEs in this way allows us to identify four parameters of interest: The Average Treatment Effect (ATE), Average Treatment effect on the Treated (ATT), Average Treatment effect on the Untreated (ATUT), and the marginal policy-relevant treatment effect (MPRTE), by calculating a weighted average of the MTE curve, with the weights depending on the specific treatment parameters. The MPRTE depends on the perturbation that defines the marginal change. We follow one of the suggestions by Carneiro et al. (2010) and define the MPRTE as a policy intervention that increases all propensity scores by an infinitesimally small fraction, e.g., by making it slightly more costly to transition into other transfers or slightly subsiding the increase in employment.[38] Lastly, we formally test the presence of

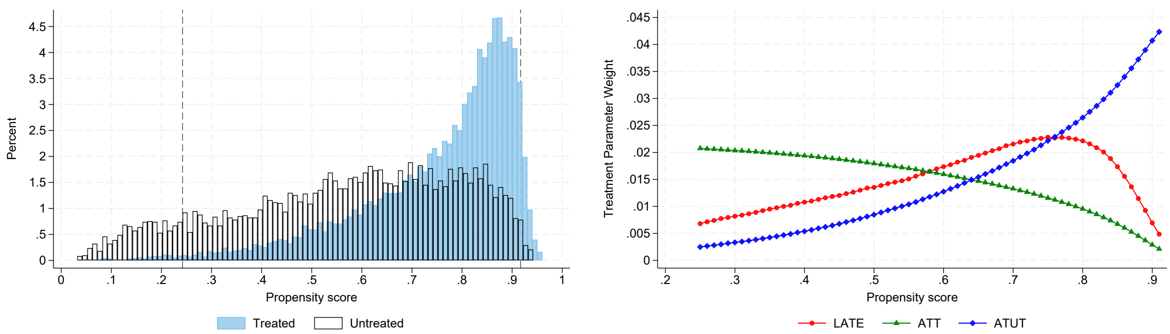
[38] In the notation of Carneiro et al. (2010), we define the MPRTE using $P_\alpha = (1+\alpha)P_0$, where P_0 is the baseline probability, $p(D_i = 1)$ and α is the marginal change in a neighborhood of the current base policy ($\alpha = 0$).

observable and unobservable heterogeneity in treatment effects by testing, respectively, the joint significance of $\beta_1 - \beta_0$ and π_l in Equations (F4) and (F5). These tests can be interpreted as tests of treatment effect heterogeneity by observables and the unobserved resistance to treatment.

The resulting propensity scores from estimating Equation (F3) are shown in Figure F.1 for the treated and untreated individuals separately. Figure F.1a shows continuous overlap in the estimated propensity scores for the two groups for scores between 0.2 and 0.9. However, a lack of overlap for scores below 0.2 is very apparent. To alleviate the issue of lack of common support, we trim 1% of the sample based on the highest and lowest propensity scores, similarly to Carneiro et al. (2011). The trimmed values are indicated by the vertical dashed gray lines in Figure F.1a. Consequently, we rescale treatment effect parameter weights to sum to 1 within this support. Figure F.1b illustrates the weights assigned to individuals in different dimensions of the propensity score used to calculate treatment effect parameters.

Figure F.1: Common Support and Treatment Parameter Weights

(a) Common Support | (b) Treatment Parameter Weights



The figure shows the estimated propensity scores for the treated and untreated individuals separately (left). Propensity scores are estimated by Equation (F6). Figure F.1a illustrates the overlap in the estimated propensity scores between the two groups for scores between 0.2 and 0.9. To alleviate the issue of lack of common support, we trim 1% of the sample with the highest/lowest estimated propensity scores, indicated by the dashed gray vertical lines. Figure F.1b illustrates the treatment parameter weights for the trimmed propensity scores.

Table F.1 presents the results of estimating Equation (F6), using 4 different specifications. All coefficients in Table F.1 are given as log odd ratios. Column (1) corresponds to Column (1) of Table 1 Panel A, but using a logit link function. Column (2) adds controls for disposable income and net wealth at ages 48-52 including squared terms. Column (3), adds interaction terms of the controls for disposable income and net wealth with the instrument, $\mathbb{1}[X_i \geq c]$. Finally, Column (4) adds control dummies for gender, being married, and education levels, as well as healthcare utilization measured four years before announcement of the reform, and occupational characteristics measured at age 55. The row "F-statistic" displays the F-statistic from a join test of $\mathbb{1}[X_i \geq c]$ and all of its interactions. In all columns, the F-statistic is greater than 700, being a sign of strong relevance of the instruments in explaining employment in the half-year of the ERA extension. All instruments are individually statistically significant at the 5 % level.

Table F.1: Estimating MTEs: First Stage

	(1) Employment	(2) Employment	(3) Employment	(4) Employment
$\mathbb{1}[X_i \geq c]$	0.793***	0.857***	2.142***	2.434***
	(0.0293)	(0.0304)	(0.221)	(0.233)
$\mathbb{1}[X_i \geq c] \times$ Income			-0.0592***	-0.0694***
			(0.0130)	(0.0137)
$\mathbb{1}[X_i \geq c] \times$ Income^2			0.000352**	0.000455**
			(0.000170)	(0.000180)
$\mathbb{1}[X_i \geq c] \times$ Net Wealth			0.227***	0.223**
			(0.0868)	(0.0904)
$\mathbb{1}[X_i \geq c] \times$ Net Wealth^2			-0.0654**	-0.0712**
			(0.0326)	(0.0340)
F-statistic	734.03	795.14	893.43	966.99
Income controls	No	Yes	Yes	Yes
Socio-economic controls	No	No	No	Yes
N	23,694	23,694	23,694	23,694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

First stage regressions estimated by Equation (F6), using varying levels of controls and interactions with the date of birth instrument, $1[X_i \geq c]$. "Income controls" refer to average net wealth and disposable incomes at ages 48-52, including squared terms. "Socio-economic controls" refer to dummies for gender, being married, and education level, as well as healthcare utilization measured over four years before announcement of the reform, and occupational characteristics measured at age 55. "F-statistic" refers to the F-statistic from a joint test of $\mathbb{1}[X_i \geq c]$ and all of its interactions. The outcome variable, employment, is in all columns measured at the half-year of the ERA extension.

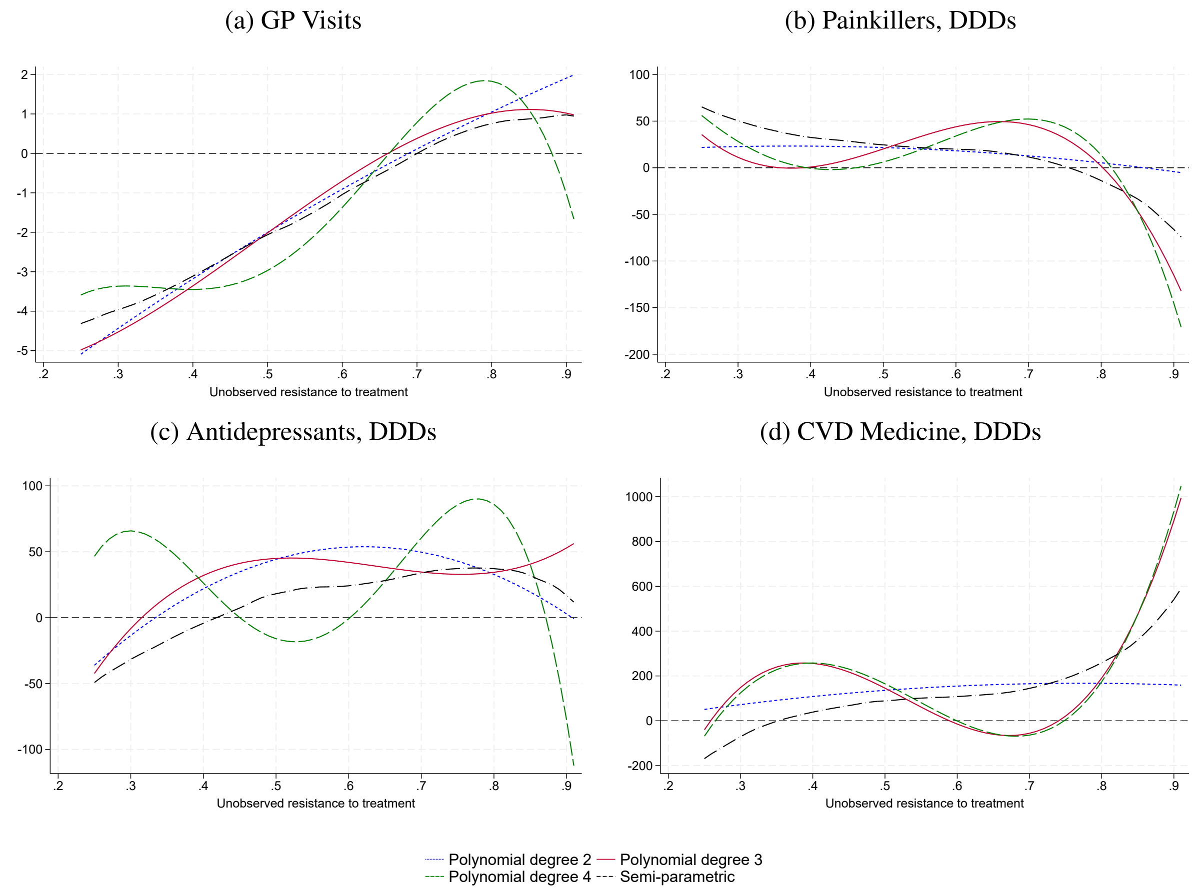
Table F.2: Marginal Treatment Effects - Different Specification Choices

	GP visits				Painkillers				Antidepressants				CVD Medicine			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
LATE	-0.0329	-0.0797	-0.0654	-0.283	-3.844	-9.706	-9.495	-14.53	12.62	15.26	16.17	2.110	91.76	130.3**	130.1*	130.7*
	(0.444)	(0.452)	(0.489)	(0.483)	(12.11)	(12.85)	(12.13)	(12.07)	(13.05)	(14.43)	(13.75)	(12.99)	(67.55)	(65.24)	(66.52)	(71.33)
ATE	-1.267*	-1.312*	-1.342*	-1.359*	15.37	9.756	9.315	13.95	28.26	30.79	28.87	13.03	135.7	172.7**	173.3*	129.7
	(0.687)	(0.675)	(0.696)	(0.750)	(20.68)	(21.63)	(21.96)	(24.21)	(28.29)	(26.53)	(27.21)	(26.96)	(92.31)	(86.26)	(90.01)	(91.99)
ATT	-2.248**	-2.229**	-2.183**	-2.199*	21.48	23.85	24.52	29.77	23.75	22.69	25.61	2.707	136.1	120.5	119.6	73.54
	(1.128)	(1.098)	(1.111)	(1.184)	(34.39)	(35.91)	(37.86)	(40.90)	(47.41)	(43.84)	(47.10)	(44.74)	(146.7)	(138.7)	(140.1)	(144.7)
ATUT	0.885	0.700	0.509	0.488	1.748	-21.42	-24.23	-20.87	38.46	48.89*	36.67	35.94	133.3	285.8**	289.7**	250.7**
	(0.806)	(0.829)	(0.841)	(0.837)	(24.01)	(25.35)	(24.34)	(24.64)	(26.77)	(28.33)	(29.19)	(26.24)	(112.7)	(119.9)	(122.9)	(126.2)
MPRTE	0.144	0.0888	0.0189	0.0127	2.876	-4.029	-5.056	-4.180	11.04	14.15**	9.686	9.621	56.86*	102.3***	103.7***	93.12**
	(0.233)	(0.237)	(0.244)	(0.246)	(6.465)	(6.530)	(5.989)	(6.169)	(6.891)	(7.037)	(7.007)	(6.564)	(33.95)	(35.61)	(37.65)	(39.12)
P-value, k(p)	0.138	0.254	0.183	0.996	0.950	0.438	0.438	1.000	0.544	0.717	0.393	0.985	0.948	0.092	0.184	0.997
P-value, $\beta_1 - \beta_0$	0.017	0.033	0.016	0.012	0.444	0.560	0.576	0.638	0.011	0.003	0.001	0.012	0.009	0.006	0.001	0.010
Specification	Polynomial, degree 2	Polynomial, degree 3	Polynomial degree 4	Semi-parametric	Polynomial, degree 2	Polynomial, degree 3	Polynomial degree 4	Semi-parametric	Polynomial, degree 2	Polynomial, degree 3	Polynomial degree 4	Semi-parametric	Polynomial, degree 2	Polynomial, degree 3	Polynomial degree 4	Semi-parametric
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456	23,456

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors in parentheses with 500 replications.

The table shows treatment effect parameter estimates under different functional form assumptions of the MTE equation, Equation (F5) for the four main outcomes. Columns (1), (5), (9), and (13) assume a second-order polynomial form of the MTE curves, Columns (2), (6), (10), and (14) assume a third-order polynomial (the main specificaiton in Table 4), Columns (3), (7), (11), (15) assume a third-order polynomial, and Columns (4), (8), (12), and (16) use semi-parametric estimation with 100 gridpoints to speed up computation. All outcomes denoted in the column headers are measured two years after implementation of the extended ERA. "Controls" included in the estimations are average wealth and disposable incomes at ages 48-52, including squared terms, and dummies for gender, being married, and education level, as well as healthcare utilization measured four years before announcement of the reform, and occupational characteristics measured at age 55. The excluded instruments are a binary variable for being born after July 1, 1954, and this binary variable interacted with the variables for disposable income and net wealth and their squared terms. The row "P-value, k(p)" refers to the P-value from a joint test of the slope coefficients of the MTE equation, Equation (F5), and "P-value, $\beta_1 - \beta_0$" refers to the P-value from a joint test of whether the treatment effects differ across observed covariates.

Figure F.2: Marginal Treatment Effect Curves - Different Specification Choices



The figure illustrates the MTE curves for different functional form assumptions, corresponding to those imposed in Table F.2. The main specification in Table 4 assumes the MTE curves are polynomials of degree 3.