The Effect of Retirement Duration on Women's Cognitive Performance*

Henrik Bergschneider
RWI Essen, and
Paderborn University

Robin Kottmann
RWI Essen, and
Paderborn University

Hendrik Schmitz
Paderborn University,
RWI Essen, and
Leibniz Science Campus Ruhr

Matthias Westphal
University of Hagen,
RWI Essen, and
Leibniz Science Campus Ruhr

Abstract

We examine the effect of women's retirement duration on various cognitive ability measures (word recall, semantic fluency, and the stroop effect) as well as on the diagnosis of cognitive disorders (CD) such as mild cognitive impairment (MCI) and dementia. As a source of exogenous variation in retirement, we utilize the 1999 reform of the German pension system which increased the early retirement age for women born after 1951 by three years form 60 to 63. Looking at women aged 63 to 67, we find statistically significant negative effects of an additional year spent in retirement on cognitive abilities and null results on CD (dementia and MCI).

Henrik Bergschneider: RWI Essen, Hohenzollernstr. 1-3, 45128 Essen, Germany, Tel.: +49 201 / 8149-340, E-mail: henrik.bergschneider@rwi-essen.de.

Robin Kottmann: RWI Essen, Hohenzollernstr. 1-3, 45128 Essen, Germany, Tel.: +49 201 / 8149-286, E-mail: robin.kottmann@rwi-essen.de.

Hendrik Schmitz: Paderborn University, Warburger Strasse 100, 33098 Paderborn, Germany, Tel.: +49 5251 603213, E-mail: hendrik.schmitz@uni-paderborn.de.

Matthias Westphal: University of Hagen, Fakultät Wirtschaftswissenschaften, 58084 Hagen, Germany, Tel.: +49 2331 987-2633, E-mail: matthias.westphal@fernuni-hagen.de.

*We thank participants of the Annual Meeting of the German Health Economics Association (DGGÖ) in Halle, the EUHEA PhD conference in Lucerne, the Westfälischer Forschungsworkshop and the RWI health department research workshop for valuable comments. Financial support from the Deutsche Forschungsgemeinschaft (DFG, project number 437564156) is gratefully acknowledged.

1 Introduction

Cognitive abilities are crucial for economic decision making (Tymula et al., 2013; Li et al., 2013). This is incorporated in different areas such as financial decisions or decisions regarding health behavior (Christelis et al., 2010; Cutler and Lleras-Muney, 2010). It is widely known that cognitive abilities decline with age (Salthouse, 2009). Moreover, ageing is associated with a higher risk of developing cognitive impairments or diseases such as mild cognitive impairment (MCI) or dementia. The ongoing demographic change with a growing elderly population, intensifies this issue, by raising the demand for long-term care due to a growing number of people suffering from cognitive diseases which ultimately puts pressure on already stressed health care systems (Rechel et al., 2013). Beyond age-related decline, life events or behavioral choices can influence cognitive abilities. In this regard, economic literature has widely studied effects with respect to human capital accumulation, e.g. education plays a key role in shaping cognitive functioning (Banks and Mazzonna, 2012; Schneeweis et al., 2014; Schmitz and Westphal, 2022). However, less attention has been paid to human capital depreciation, particularly how behavioral choices and decisions throughout life shape cognitive decline (McFadden, 2008). As the number of people with cognitive impairments rises, it becomes increasingly important to understand these factors with the aim of preserving cognitive abilities in older ages.

We investigate one major period of time in the life of an individual that may have an impact on the decline in abilities: the period of retirement. We are not the first to study this research question. Previous literature so far suggests that retirement negatively affects cognitive abilities (Rohwedder and Willis, 2010; Bonsang et al., 2012). Other studies confirm these findings and additionally show that length of retirement increases this negative effect (Celidoni et al., 2017; Mazzonna and Peracchi, 2012, 2017). Furthermore, there is heterogeneity with respect to type of former work (Mazzonna and Peracchi, 2017; Coe et al., 2012); and also with respect to gender (Atalay et al., 2019; Mazzonna and Peracchi, 2012; Mosca and Wright, 2018). Dufouil et al. (2014), Grotz et al. (2015) and Sundström et al. (2020) find an association between late retirement age and risk of dementia, but without claiming causality.[1] Schmitz and Westphal (2021) find negative average effects of retirement on cognitive abilities for individuals aged 50 to 75 up to 10 years after retirement entry: early retirees only experience small or no negative effects while late retirees lose up to 20% of their cognitive abilities. Rohwedder and Willis (2010) conclude that retirees, unlike workers, less face mental challenging tasks and are not in a cognitively stimulating work environment.

[1]Moreover, our paper relates to the literature on the impact of old-age retirement on various health outcomes. In this regard, there are studies that utilize diagnoses and drug prescriptions from administrative data of public health insurances to indicate health outcomes such as mental health, musculoskeletal diseases, or obesity (Barschkett et al., 2022; Kuusi et al., 2020; Nielsen, 2019; Hagen, 2017; Horner and Cullen, 2016).

Women face different employment situations throughout their lives compared to men, which could potentially influence the effects of retirement on cognitive abilities. However, the role of women has changed in recent decades. In Germany, womens' propensity to work is increasing compared to the past (Federal Statistical Office, 2024). This could be due to a better reconciliation of family and working life. There are four papers that explicitly investigate effects of time spent in retirement on cognitive abilities for women or distinguish effects by gender (Mazzonna and Peracchi, 2012, 2017; Mosca and Wright, 2018; Atalay et al., 2019). Mazzonna and Peracchi (2012, 2017) use cross-country variations in ERA and find negative effects for women. Mosca and Wright (2018) use the abolition of the Irish marriage-bar and find zero effects. Atalay et al. (2019) also find zero effects for Australian women while effects for men are present. We contribute to these works in two ways: First, we exploit an inner-country reform in Germany that explicitly affected women's decisions to retire. Therefore, we can make sure that potential effects are not caused by variations from different retirement systems of different countries but only rely on exogenous variations of women's retirement duration. By doing that, we add new evidence of a negative effect on cognition measures that indicate fluid intelligence while we find no effect on crystallized intelligence. Second, to our knowledge, we are the first to investigate medical outcomes like MCI and dementia, but find zero effects.

We utilize data from different sources. First, we use the German National Cohort (NAKO), an ongoing examination that investigates medical evaluations, as well as sociodemographic and socioeconomic factors of individuals aged between 20 and 69. We use the NAKO to estimate effects on neuropsychological cognition measures. These include word recall, semantic fluency, and the stroop effect. Second, we use a large and nationally representative sample of data from a German health insurer that covers about 10% of the German population. The data contains diagnoses and demographic characteristics of the individuals, such as sex and birth year. We utilize this information to explore diagnoses of cognitive disorders (CD) such as dementia and MCI.

The decision to retire constitutes a personal choice and may also be influenced by a deteriorating mental health. To address this, our identification strategy exploits the 1999 German pension reform, which raised the early retirement age (ERA) for women born after 1951 by three years, in order to capture exogenous variation in retirement. More precisely, this paper focuses on the impact of retirement duration, measured in years, on the cognitive functioning of women. Thus, we estimate local average treatment effects (LATE) for women born between 1949 and 1954 from ERA age 63 to 67. We find that an additional year spent in retirement reduces the delayed word recall by 0.24 (3 % of the sample mean, 11 % of the standard deviation) and the sum of recalled words by 0.53 (2 % of the sample means, 11 % of the standard deviation). In addition, the results show that the response time difference included in the stroop effect worsens by 1.4 seconds which is 6 % of the sample mean and 11 % of a standard deviation. Although the results indicate that a longer time spent in retirement is associated with a decline in women's cognitive abilities, we do not observe significant effects on medical outcomes related to cognitive health. This is unsurprising, as symptoms of MCI and dementia commonly manifest after the age of 70 (OECD, 2018). It is up to future research to address this and investigate whether the effects become noticeable at later ages.

Our findings suggest that policy interventions aimed at postponing the transition into retirement, with the goal of strengthening the workforce and the social security system, may yield positive outcomes for individuals' mental well-being. One possible mechanism for these effects is the cognitive engagement maintained through continued employment. This is further supported by the monetary incentives that encourage individuals to preserve cognitive functioning. However, our results do not demonstrate external validity of these effects for population groups other than women born between 1949 and 1954. Nonetheless, it is conceivable that these findings could be applicable to other population groups.

The paper is structured as follows. Section 2 describes backgrounds of the pension reform that we use as an instrument for our empirical approach. Section 3 presents the data and the variables used. The empirical analysis includes the estimation strategy as well as the results and is outlined in Section 4. Section 5 concludes the analysis.

2 Pension Reform Background

The German pension system differentiates between an official retirement age (ORA) and an early retirement age (ERA). Individuals born before 1947 could officially retire at the age of 65. For those born in 1947 and later, the ORA is raised step-wise to 67 by 2031 (birth cohort 1964 and later). Moreover, the pension system grants the possibility to retire early. However, individuals face permanent deductions in their pension benefits when applying to the ERA. The German pension reform of 1999 changed the possibility to retire early for women born in 1952 or later. Before the establishment of the reform, the German pension system offered women the opportunity to retire early at the age of 60 while the ORA was 65. This regulation was commonly known as the "pension for women". The precondition to enter early retirement included a coverage by the statutory pension insurance for at least 15 years as well as contributions to the pension fund for 10 years after their 40th birthday[2]. Women who chose to retire early face a permanent pension benefit deduction of 0.3 % for each month they undercut the ORA. This deduction regulation was introduced to incentivize a postponing of the retirement. Before the pension reform came into force, women who retired at the age of 60 had faced a permanent deduction from their pension benefits, which accumulated to 18%.

[2] Geyer and Welteke (2021) show that 60 percent of women born in 1951 were eligible for the pension for women.

The reform of 1999 abolished the pension for women with the aim to extend women's participation in the labor market. In particular, the ERA for women born in 1952 or after was adjusted to 63 whereas women born before January 1st, 1952, were still able to retire by the age of 60. Consequently, this reform implied a strict demarcation regarding the ERA for different birth cohorts. We exploit this reform as an exogenous variation of retirement duration. This has previously been done by other researchers. Geyer and Welteke (2021) examine effects of the reform on different labor market outcomes. The authors find that the pension reform had a statistically significant effect of 13.5 percentage points on the employment rate of women aged between the old and new ERA. This equals a 30% increase compared to the pre-reform mean. Geyer et al. (2020) investigate labor market effects of the reform. They find that the reform led to a significant drop in retirement rates for the affected birth cohorts. Women aged between the old and new ERA are 16 percentage points less likely to be retired which equals a 50% drop compared to pre-reform means. Simultaneously, the reform had a positive impact of 8.4 percentage points on employment rates. Women born in years after 1951 belong to the work force for a longer duration compared to women born in 1951 or before. The empirical results of these studies indicate that the pension reform had significant effects on the labor market outcomes of affected individuals, such as retirement and employment. We build on this by also using the reform as an instrument to model exogenous variations of years spent in retirement.

Moreover, other studies analyze effects of the German pension reform on various outcomes. Etgeton et al. (2023) find that households decrease savings rates and increase leisure spending in response to the reform. These effects are driven by households with married women. Fürstenau et al. (2023) show the positive effect of the reform on participation in on-the-job training with a sharp RDD design indicating positive employment effects. The pension reform also has been used to show effects on health care outcomes (Barschkett et al., 2022; Geyer et al., 2023). Barschkett et al. (2022) find that increasing the early retirement age negatively affects health outcomes in terms of mental health, musculoskeletal diseases and obesity. Furthermore, in a follow-up study, the authors show positive effects on health care costs (Geyer et al., 2023).

3 Data

We use data from two sources. First, we introduce the German National Cohort (NAKO). It is an examination that covers information on medical inspections as well as sociodemographic and socioeconomic factors of individuals living in Germany (Peters et al., 2022). Second, we use administrative health records for nine million individuals (over ten percent of the German population) insured with one of Germany's largest public health insurers (Grobe and Szecsenyi, 2023).

3.1 NAKO - The German National Cohort

The NAKO is the largest epidemiological examination in Germany to date, that involves the interdisciplinary collaboration of 27 German scientific institutions, including 15 universities, 4 Helmholtz health centres, 4 institutes of the Leibniz Association and 4 other national research institutes. The aim of the NAKO is to investigate causes and the development of major chronic diseases. Beginning in 2014, the NAKO has invited about 200,000 individuals aged between 20 and 69 to one of their 18 different study centers throughout whole Germany covering both rural and industrial areas. It is an ongoing, long-term population study. The first investigation of the participants, the Level-1 baseline examination, took place between 2014 and 2019 and assessed basic medical and socioeconomic factors. A share of 20% of the participants undergo further medical examinations (Peters et al., 2022).

For our analysis we use the baseline examination of the NAKO. It provides information on both measures of cognitive abilities of individuals and information on retirement and labor force participation. As we predominantly investigate retirement induced by an advancing age and not by premature health related reasons, we restrict our sample to individuals that have been retired for a maximum duration of 10 years. We only consider women born between 1949 and 1954, that is, women born around the reform eligibility birth year of 1952. We observe this group of women at ages that are associated with retirement and expect them to perceive a variation regarding retirement duration induced by the pension reform. The NAKO data does not provide information on the exact birth year of an individual. Instead we calculate the birth year based on the age at the examination and the respective examination calendar year. This results in a possible misspecification of the 1952 cohort regarding their treatment by the reform affecting our first stage. We address this limitation by performing a robustness check where we drop the 1952 cohort and see if the results are stable.

Table 1 shows the number of observations by age and calculated birth cohort. Since the baseline study of the NAKO has been conducted between the years 2014 and 2019, we are able to observe six different age groups for each of the birth cohorts ranging from 1949 to 1954. For age 65, the sample provides observations for each of the six birth cohorts. If the birth cohorts are divided into a group affected by the reform (1952 - 1954) and a group not affected by the reform (1949 - 1951), this overlap extends to the ages 63 to 67. We can therefore observe labor force participation at five different ages in both the birth cohorts of women affected by the reform and the birth cohorts of women not affected by the reform. We apply this to the sample and only include women aged 63-67. This age interval leaves 5,120 observations in the unaffected group and 4,718 observations in the affected group.

Table 1: Number of observations by age and calculated birth year in NAKO

	Calc. birth year						Reform	
Age	1949	1950	1951	1952	1953	1954	Not affected	Affected
60	0	0	0	0	0	109	0	109
61	0	0	0	0	118	402	0	520
62	0	0	0	100	453	563	0	1,116
63	0	0	109	507	629	558	109	1,694
64	0	125	496	663	570	354	621	1,587
65	116	498	685	570	358	66	1,299	994
66	488	597	572	317	63	0	1,657	380
67	589	504	341	63	0	0	1,434	63
68	458	272	63	0	0	0	793	0
69	223	49	0	0	0	0	272	0
70	25	0	0	0	0	0	25	0
Observations								
Total	1,899	2,045	2,266	2,220	2,191	2,052	6,210	6,463
Sample	1,193	1,724	2,203	2,120	1,620	978	5,120	4,718

Note: This table displays the number of observations by age and calculated year of birth in the NAKO dataset.

3.2 Administrative health insurance data

Our second dataset is a panel dataset consisting of administrative health records from 2012 to 2019. Figure A3 displays the geographical distribution of insured individuals across Germany. In addition to all inpatient and outpatient health records, diagnoses, prescriptions, and medical billings, it includes individual-level characteristics such as birth year, date as well as type of retirement and sex. The sample we take from this dataset is constructed to be as similar as possible to our NAKO sample while still utilizing its size and time horizon. We consider individuals born from 1949 to 1954 aged 63 to 67 years old. Table A3 shows descriptive statistics for relevant variables. Figure A4 shows that we are able to observe individuals of all relevant ages in cohorts before and after the reform. We exclude individuals with other reasons for retirement than age, for example due to a health related inability to work. This results in around 2 Million individual-years between the years 2012 and 2019.

3.3 Outcome variables

The "ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning to overcome obstacles by taking thought" are defined as cognitive abilities (American Psychological Association, 1995). The sum of these abilities is referred to as intelligence. Psychologists differentiate between two sorts of intelligence – fluid and crystallized intelligence. While fluid intelligence measures the innate cognitive ability capturing memory and processing speed, crystallized intelligence measures abilities and knowledge that are developed during the lifetime due to education and experience (Rohwedder and Willis, 2010).

The NAKO employs a variety of cognitive tests, referred to as the Level 1 cognitive test battery. This battery consists of six tasks. We use these tasks to generate four cognitive scores that represent our outcome measures[3]. We use delayed word recall, the sum of words recalled, semantic fluency and the stroop effect. These are neuropsychological tests, that aim on assessing verbal episodic memory, processing speed or executive control[4].

As an addition to the investigation of cognitive abilities, we further provide insights of effects on objective health outcomes associated with cognitive impairment by using health insurance claims data. We include MCI acting as preliminary diagnosis to dementia as explained in German coding guidelines by DGN e. V. and DGPPN e. V. (2023). We combine it with diagnoses of dementia to a single outcome measure of cognitive decline[5]. Our administrative dataset is on individual-year level with our outcome variable being equal to 1 if an individual is diagnosed in the current year.

Word recall

The first cognitive measure we employ is the delayed word recall which is implemented as the final task in the NAKO test battery. Prior to its measuring, participants must undergo the previous evaluations. In step 2 of the test battery, participants were asked to undertake an immediate recall test. A digitally recorded list of 12 different German nouns was presented to the participants. Each word is presented in a two-second period. Subsequently, participants were asked to recall as many words as possible. This procedure was repeated twice. Following this, participants were informed that they would undergo a delayed recall test at the end of the test battery. This delayed word recall test only relies on their memory. The count of accurately recalled words in the delayed recall test gives our first measure of cognitive abilities in the application. In addition, we use the sum of words recalled, which is calculated by adding the number of words from the two immediate word recall tasks and the delayed word recall task. This means that this score can range between 0 and 36 words. Measures of word recall are used in many economic papers that analyze cognitive abilities, for instance Rohwedder and Willis (2010), Coe et al. (2012), Mazzonna and Peracchi (2012), Celidoni et al. (2017), Schmitz and Westphal (2021) or Freise et al. (2022). Word recall is associated as a measure of fluid intelligence.

[3]A detailed description about the test battery and the different cognitive measures can be found in Kleinedamm et al. (2022).

[4]See Figure A1 and Table A2 for summary statistics of outcome variables and individual characteristics.

[5]See Table A1 for detailed ICD-10-GM Codes used in our analysis.

Semantic fluency

Our second measure implies semantic fluency. Participants were assigned the task of enumerating as many animal names as they could within a one-minute time frame. The number of animal names mentioned provides the measure of semantic fluency. Schiele and Schmitz (2023) use the number of animals mentioned as a measure of verbal fluency to examine the effect of health shocks on cognitive abilities. Freise et al. (2022) use verbal fluency to investigate effects of late-career unemployment on cognition. Verbal or semantic fluency is particularly present in research on neuropsychology, specifically thematizing Alzheimer's disease or dementia. It is subject of various works, all of which are outside the economic literature, e.g. Murphy et al. (2006), Mueller et al. (2015) and Nikolai et al. (2018). Semantic fluency is considered as a combined measure of fluid and crystallized intelligence as it acquires both knowledge obtained due to education and the property of processing speed.

Stroop effect

The fourth measure of cognitive abilities employed in our study is the stroop effect. This effect quantifies the temporal difference between two stroop task tests that were carried out in the test battery previously. In the initial stroop test, participants were instructed to name the colors of 36 differently colored boxes (red, green, blue, yellow). In case of an error, participants were interrupted and corrected. The score is defined as the time to name the colors of all the boxes. For the subsequent stroop test, participants were presented 36 printed color names where the color of the letters did not match the color names. Participants were asked to identify the color of the letters (e.g., if the word "RED" was printed in green, the correct response would be "green"). Again, participants were interrupted and corrected in case of errors. The time to name all 36 colors was recorded. The Stroop effect measure was calculated as the difference in time between the second test and the first test. A large difference between the two test measures indicates poorer cognitive ability. We recode the stroop effect by multiplying it with the factor -1 to make the direction of potentials effects consistent to the other cognitive measures. The stroop effect is a measure of fluid intelligence and refers to response inhibition. This is the ability to consciously inhibit dominant or automatic responses (Miyake et al., 2000). The time difference between a baseline stroop task and a more complex stroop task shows whether a person has better control over response inhibition. Different works and meta-analysis on the association of aging, mental diseases and cognitive functioning are relying on the stroop test, for instance Spieler et al. (1996), West and Alain (2000) or Ben-David et al. (2014). These works are predominantly located in the field of neuroscience.

Administrative data: Diagnoses of cognitive disorders

We use ICD-10-GM codes to identify dementia and mild cognitive impairment in our data as diagnoses of cognitive disease (CD). According to OECD (2018), in the EU in 2015 only 2 percent of women aged 60 and 64 and 3 % aged between 65 and 69 suffered from dementia. Since individuals born around the reform cutoff are 67 years old in 2019 we did not only rely on the use of dementia for our analysis. Instead we added the preliminary diagnosis of MCI and combined it with dementia, resulting in our pooled medical outcome for diagnoses of CD[6]. Petersen et al. (1999) define MCI as follows: "Subjects with a mild cognitive impairment (MCI) have a memory impairment beyond that expected for age and education yet are not demented". A meta-analysis by Bai et al. (2022) estimated the global prevalence of individuals in our relevant age group between 60 and 69 years old to be 11.5 %. A recent medical guideline for German physicians on dementia highlights ICD-10-GM F06.7 (Mild cognitive impairment) as an unspecific diagnosis of cognitive impairment without the severity of dementia DGN e. V. and DGPPN e. V. (2023). A German ICD-10 classification manual by the Zentralinstitut kassenärztliche Versorgung (2024) recommends to use MCI as a diagnosis if a patient shows symptoms of cognitive impairment such as mild forgetfulness without fulfilling the criteria for dementia or having other mental or behavioral disorders (ICD-10 F10 to F99).

3.4 Retirement

To make generally valid assumptions on the retirement status of women in the NAKO we use information on labor force participation that are available in the dataset. The labor force status differentiates between three categories: employed, unemployed and not in the labor force. We consider women as retired if they declare their status as not being part of the labor force and state a specific retirement age. The probability of women not participating in the labor force rises strongly from the age of 60 onwards, providing graphical evidence for our definition of being retired (Appendix Figure A2). By implication we define women as not being retired who make a statement about their retirement age but are part of the labor force because they work part- or full-time. We calculate the retirement duration as the difference of age by the time of the examination and stated age of retirement. If stated age and retirement age are equal, we define women as being retired for one year. Women who are not yet retired take a retirement duration of 0.

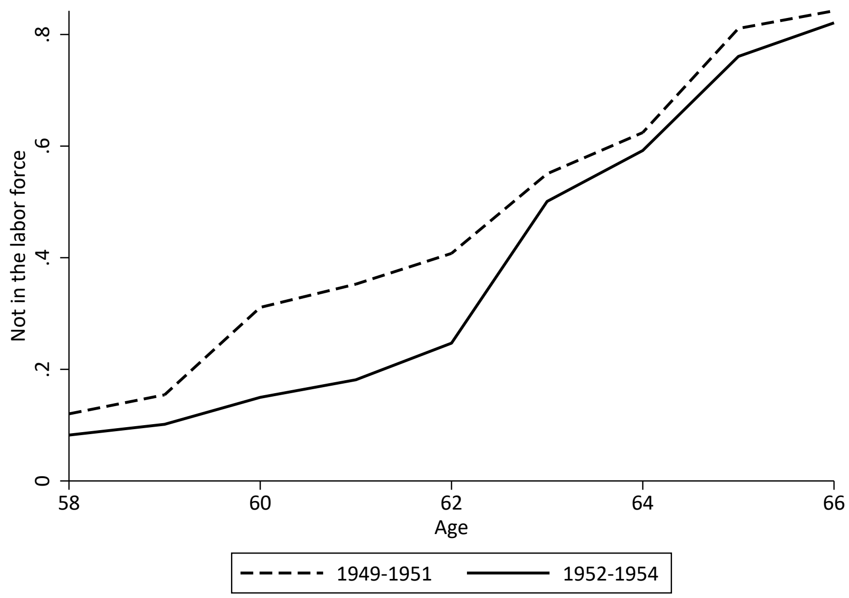
We present graphical evidence that the reform has had an impact on the decision to retire for the birth cohorts we inspect. As the reform raised the ERA from 60 to 63, differences in labor force participation at these ages are likely. Table 1 shows that the earliest age observed for cohorts affected and unaffected by the reform is 63. As 63 is the post-reform ERA, meaning that any birth cohort can retire, there are no striking differences in retirement status between cohorts. We investigate whether the pension reform had an impact on the retirement age of women by creating an artificial panel of the dataset observing the

[6]See Table A3 for summary statistics and Figure A6 for CD diagnosis by age.

labor force status for previous years based on the respondent's reported retirement age. We count back the age accordingly and calculate average labor force participation of the two groups (1949-1951 and 1952-1954) by age. Figure 1 shows the differences in labor force status between cohorts. We follow Geyer and Welteke (2021) who do this with administrative data from public pension insurance accounts. These data allow the authors to observe the actual retirement status of women at different ages. With our artificial panel approach we find similar results. The probability of nonparticipation in the labor force starts rising strongly for the birth years 1949-1951 by the age of 60. The birth years 1952-1954 only register a sharp rise by the age of 63. The space between these lines defines the exogenous variation of non-participation in the labor force that is determined by the pension reform. We use this gap to estimate effects of retirement duration on cognitive abilities.

In addition, we confirm the effect of the reform using our administrative dataset. The dataset contains information on year of birth and the exact date of retirement. Figure A5 shows the share of retirement by age and cohort.

Figure 1: Differences in labor force status between birth cohorts



Note: This figure shows the labor force status of women by their age and year of birth based on an artificial panel constructed with the NAKO dataset. The solid line represents women affected by the reform and the dotted line women not affected by the reform. Information of the examination year, the age of an individual in the examination year and her age at the year of retirement are provided by the NAKO.

4 Empirical Analysis

After introducing the estimation strategy, we present our results on the relationship between retirement duration and cognition measures examined in the NAKO. We then report findings of effects on medical indicators of CD using insurance claims data.

4.1 Estimation strategy

By implementing a two-stage-least-squares model (2SLS), we estimate the effect of retirement duration on cognitive abilities. We exploit the pension reform of 1999 which serves as a natural experiment. We use the reform as exogenous variation in retirement duration of women in the sample. This approach follows Geyer and Welteke (2021) and Geyer et al. (2020) who investigate labor market effects of the reform giving evidence to the relevance of our instrument. Equation 1 shows the baseline OLS specification of our estimation strategy.

$$Y_{i,a,c} = X_i'\beta + \rho YearsRetired_i + \gamma_a + \lambda_c + \epsilon_{i,a,c} \tag{1}$$

$Y_{i,a,c}$ is our outcome variable denoting different measures of cognitive abilities, namely delayed word recall, sum of words recalled, semantic fluency and the stroop effect. $YearsRetired_i$ is the retirement duration of an individual measured in years. ρ gives the effect retirement duration has on cognitive abilities. Furthermore, we control for study center (λ_c) and age (γ_a) fixed effects. X_i represents the set of covariates used. Since our measures of cognitive abilities are highly correlated to the proficiency of speaking German, we control for birthplace[7]. We assume that an individual that is born in a foreign country might not have the same level of German fluency as someone born in Germany and would perform worse in cognition tests compared to a native speaker. To account for this, we include a dummy variable indicating whether the country of birth is Germany or not. We furthermore control for dummies indicating the highest level of education according to the International Standard Classification of Education (ISCED). In our 2SLS approach, we instrument retirement duration with a dummy indicating whether an individual was born in 1952 or after.

[7] See Kleinedamm et al. (2022)

4.2 Assumptions

In order to interpret ρ as the LATE, different assumptions on the instrument need to be fulfilled. First, the exclusion restriction needs to hold. It can be assumed that the only impact the reform has on cognitive abilities of individuals is due to its effect on retirement duration. Second, the pension reform should have a significant effect on the number of retirement years. This property can be investigated in our first stage results. Third, there are no defiers, i.e. women who decide to retire at 60 only because they are affected by the reform and would not do so otherwise. This is highly unlikely. Assuming these conditions are met, ρ gives the effect of retirement duration on cognitive abilities for the complier group, women who change the retirement date due to their inclusion in the pension reform.

Potential threats to the identification strategy

There are three potential threats to our identification strategy posed by the nature of our data. As mentioned above, we are not able to observe the exact birth year of an individual in the NAKO dataset. We calculate the birth year from the difference between age and interview year. This raises the potential problem of incorrect specified birth years regarding treatment by the reform. If a person becomes older in the same interview year after the interview has been completed, we incorrectly specify her true birth year. The 1952 birth cohort is particularly exposed to this problem, as it is the only birth cohort that is potentially incorrect with respect to being eligible of the reform. We address this by excluding individuals born in 1952 from our baseline regression as a robustness check.

A second limitation comes from possible cohort effects driving our results. Birth cohorts differ regarding their initial conditions that are shaped by environment, life circumstances or school quality. Case and Paxson (2010) show that health differences between cohorts in early life are significantly associated with cognitive performance in older ages. Moreover, the "Flynn Effect" explains that later born cohorts have better cognitive skills expressed by the IQ than earlier born cohorts (Flynn, 1987). Mazzonna and Peracchi (2012) additionally argue that cohort effects could also be driven by larger mortality rates in older cohorts and use the longitudinal structure of the SHARE to prove robustness. In contrast to our analysis, the authors look on a wide range of birth years from 1934 to 1956. In their first-differencing approach, they define birth year groups of five years and thus can demonstrate robustness to possible effect heterogeneities. As our samples in combination with our empirical approach do not allow for a proper separation of age and cohort effects. We address this issue by following Salthouse (2013) who uses a similar structured dataset as we do and shows that age effects within a single cohort correspond to age effects examined at different cohorts. Nevertheless, as we only include six different birth years, we are confident that potential cohort heterogeneities are only a minor threat to our estimation strategy. Previous literature usually looks on a broader set of birth years when investigating cohort effects (Gerstorf et al., 2011; Zelinski and Kennison, 2007; Karlsson et al., 2015).

The third potential concern arises from the distribution of observations in datasets regarding age and treatment by the reform as shown in Table 1 and Figure A4. Due to the nature of our data, in wave 2014 we only observe women who remain unaffected by the reform (birth years 1949 – 1951). On the other hand, in wave 2019 we only observe women who are affected by the reform (birth years 1952 – 1954). Assuming that wave effects are present, we could potentially misspecify an effect of retirement duration. We therefore check for robustness against wave effects by excluding examination periods 2014 and 2019 from the baseline regression.

4.3 Main results

Columns 1 to 4 in Table 2 present the regression results for the different outcome measures of cognitive abilities specified with Equation 1. OLS results, first stage estimates as well as 2SLS results with full controls are reported.

OLS results show significant correlations of retirement duration with cognitive abilities measured by word recall. An additional year in retirement is associated with a decrease in delayed word recall by 0.017 words and a decrease in the sum of words recalled by 0.047. Estimates of semantic fluency and stroop effect show negative correlations that are not statistically significant. Nevertheless, since retirement duration is endogeneous, OLS results do not offer insights on potential effects on cognitive abilities.

The 2SLS results state evidence of an effect of retirement duration on cognitive abilities. We can show this for women who change their retirement entry because of the pension reform. The first stage is significant on 1% level giving credibility to the strength of our instrument. Looking on the 2SLS results reveals that the direction of the second stage effects correspond with the OLS results. We find effects that indicate a decline in cognitive abilities due to an increasing retirement duration. An additional year in retirement lowers the delayed word recall by 0.24 words which is 3% of the sample mean and 11% of a SD. This effect is statistically significant on 5% level. The result for sum of words recalled follows a similar pattern. Here, an additional year in retirement causes a decrease of 0.53 words which is 2% of the sample mean and 11% of a SD. This is significant on 5% level. Results with respect to the stroop effect confirm these findings. An additional year in retirement increases the time difference between the two stroop tests by 1.43 seconds which is 6% of the sample mean and 11% of a SD. This is significant on 5% level. Even though estimates for semantic fluency are not statistically significant they correspond to the other estimates and indicate a negative relationship.

Table 2: Main results - OLS and 2SLS

	1949-1951 vs. 1952-1954				
	Delayed word recall (1)	Sum of words recalled (2)	Semantic fluency (3)	Stroop effect (4)	Diagnosis of CD (5)
OLS					
Years retired	-0.017*	-0.047**	-0.043	-0.075	0.00017***
	(0.010)	(0.021)	(0.027)	(0.060)	(0.00004)
First stage					
Birth cohort $\geq$ 1952	-0.608***	-0.608***	-0.602***	-0.601***	-0.880***
	(0.066)	(0.066)	(0.066)	(0.066)	(0.002)
2SLS					
Years retired	-0.239**	-0.529**	-0.411	-1.428**	0.00014
	(0.107)	(0.225)	(0.280)	(0.640)	(0.00013)
Controls	Yes	Yes	Yes	Yes	No
Study center FE	Yes	Yes	Yes	Yes	No
Age FE	Yes	Yes	Yes	Yes	Yes
Mean of DV	7.48	22.93	24.05	-25.66	0.00628
	(2.23)	(4.78)	(6.33)	(13.39)	(0.07899)
Observations	8,668	8,664	8,757	8,659	1,853,104

Note: This table shows coefficients for the outcome variables of cognitive abilities and cognitive disease. The first row shows the estimated coefficients from an OLS regression based on Equation 1. The second row shows the results from the first stage and the third row from the corresponding 2SLS estimation. The regressions in columns 1 to 4 include a dummy for birthplace Germany and education dummies according to the ISCED standards. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Assuming that our measures for cognitive abilities are internally valid, the results indicate that an additional year in retirement causes a significant decline in the cognitive functioning of women. This contradicts results of Atalay et al. (2019) and Mosca and Wright (2018) who find zero effects for women in Australia and Ireland. However, our findings correspond to Mazzonna and Peracchi (2012) who find significant negative effects of an additional year spent in retirement on delayed word recall when controlling for education. The effect size is lower compared to our findings and amounts to less than 2% of their sample mean. Contrary to our analysis, they also find significant negative effects on verbal fluency. The authors later confirm their findings in an extended analysis (Mazzonna and Peracchi, 2017). Moreover, our findings are in line with literature in neuropsychological research that portrays different patterns in different fields of intelligence during the lifespan of an individual Rohwedder and Willis (2010); McArdle et al. (2002). Semantic fluency, expressed by the enumeration of animal names, corresponds to a form of crystallized intelligence whereas word recall and stroop effect show fluid intelligence. Since fluid intelligence is exposed to a stronger decline during the lifespan compared to crystallized intelligence our findings are reasonable.

In addition to our main results on measures of cognitive abilities we utilize the same empirical setup but a different dataset to estimate a LATE of additional years of retirement on the risk of a diagnosis of CD. We build this analysis on the same assumptions as before. The estimation formula is similar to Equation 1, but we exclude control variables and study center fixed effects since our sample is representative for the German population.

Column 5 in Table 2 displays the results of the first stage and 2SLS regressions estimating the effect of additional years of retirement on the risk of a diagnosis of CD. The first stage using our administrative dataset confirms a strong relationship between the reform and the duration of retirement of individuals. The 2SLS results show the estimated LATE of years in retirement on the risk of being diagnosed with CD for individuals aged between 63 and 67. Overall, the share of observations with a relevant diagnosis is 0.6 %. The estimated effect is a nonsignificant increase of the risk of being diagnosed by 0.014 percentage points by year of retirement. This equals 2 % of the sample's mean risk of a CD diagnosis or 0.7 % of a SD. The effect is not significant at any conventional levels of significance, but the relative size of the effect is in line with the estimated effects in columns 1 to 4. It indicates a decrease in cognitive abilities by years of retirement.

4.4 Robustness checks

We conduct several robustness checks. First we investigate a potential misspecification of treated cohorts in the NAKO. Then, we prove robustness against cohort heterogeneities. Lastly, we check for wave effects.

4.4.1 Misspecification of birth year

As discussed earlier, the NAKO dataset does not provide information on individuals' birth years. Therefore, we calculate the birth year using the examination year and the respondent's reported age. This method introduces the risk of incorrectly classifying women born in 1952 as being affected by the reform, even though they may not be. As a result, this could lead to an underestimation of the first-stage effect in absolute terms, as well as an underestimation of the reduced-form effects. Specifically, cohorts that may have already retired by age 60—and potentially show lower cognitive performance due to more years spent in retirement—could mistakenly be treated as if they were affected by the reform.

To address this issue, we exclude women born in 1952 from our sample. This approach allows us to more accurately distinguish between treated and untreated individuals.

However, a significant drawback is the loss of statistical power, as excluding this cohort reduces the sample size by approximately 20 percent and excludes some individuals who may have been correctly classified. To assess the robustness of our results given this limitation, we re-estimate our models with the revised sample and present the findings in Table 3.

Table 3: Excluding 1952 cohort - OLS and 2SLS

	1949-1951 vs. 1953/1954			
	Delayed word recall (1)	Sum of words recalled (2)	Semantic fluency (3)	Stroop effect (4)
OLS				
Years retired	-0.020*	-0.052**	-0.051*	-0.077
	(0.011)	(0.024)	(0.030)	(0.068)
First stage				
Birth cohort $\geq$ 1952	-0.736***	-0.734***	-0.743***	-0.734***
	(0.085)	(0.085)	(0.084)	(0.085)
2SLS				
Years retired	-0.121	-0.332	-0.468	-1.787**
	(0.114)	(0.241)	(0.304)	(0.714)
Controls	Yes	Yes	Yes	Yes
Study center FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Observations	6,823	6,821	6,897	6,816

Note: This table shows coefficients for the outcome variables of cognitive abilities. Compared to Table 2 individuals born in 1952 are excluded here as a robustness check for the calculated year of birth. The first row shows the estimated coefficients from an OLS regression based on Equation 1. The second row shows the results from the first stage and the third row from the corresponding 2SLS estimation. The regressions include a dummy for birthplace Germany and education dummies according to the ISCED standards. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

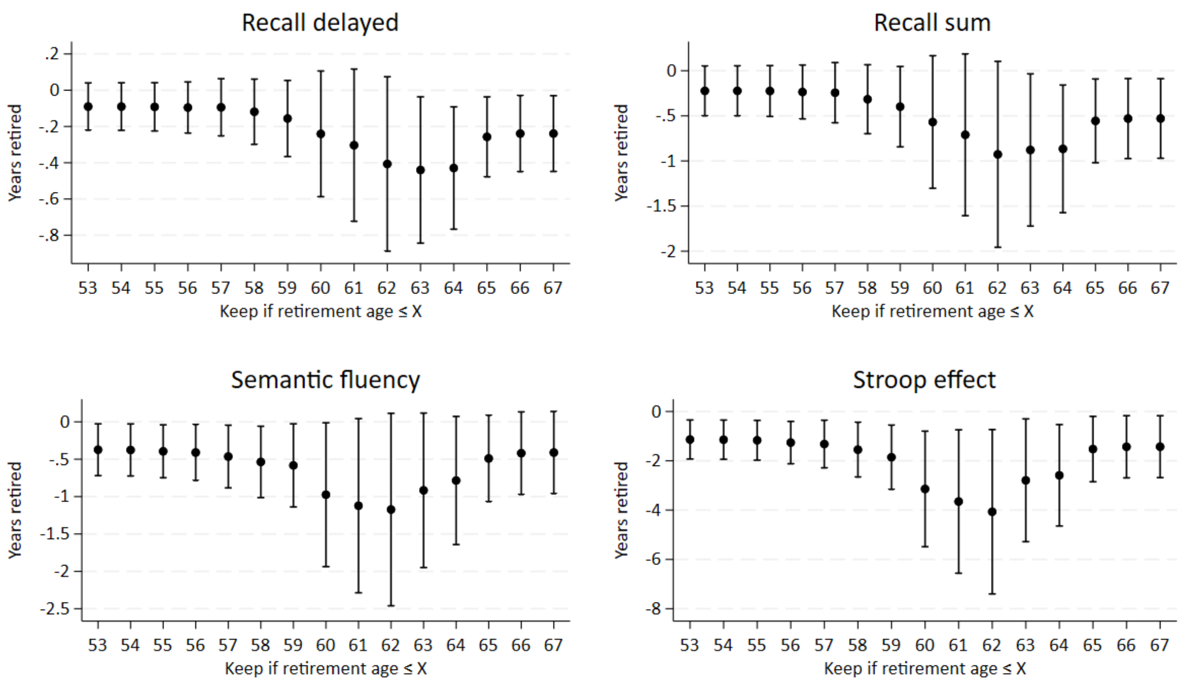
First, we examine the OLS estimates, which are consistent with the findings in Table **??**. The first-stage effects remain highly significant, and the coefficients are larger compared to those in the full sample. In the second-stage estimates, the effects on the two-word recall measures are slightly attenuated and no longer statistically significant, relative to the previous results. However, the effects on the Stroop effect are larger and statistically significant. Standard errors remain unchanged across both stages. Although the estimates for some outcomes have lost statistical significance, the negative effects persist and do not suggest a definitive zero effect.

Since the exclusion also affects correctly specified individuals, we provide additional insights into the coefficients when excluding the 1952 cohort. We estimate the coefficients from Equation 1 across 15 subsamples, gradually increasing the sample size by including individuals as the retirement age increases. The aim is to identify the threshold at which the coefficients become statistically significant. Given that the early retirement age (ERA) differs between the 1951 and 1952 cohorts, we expect that women born in 1952 retire later than those born in 1951. Thus, we hypothesize that the coefficients become more accurate as we include later retirement ages, which should predominantly capture correctly specified women from the 1952 cohort.

Unlike in Table 3, this analysis includes women who have not yet retired. Figure 2 presents the evolution of the coefficients across the subsamples. The pattern is consistent across all cognitive measures: the coefficients for years retired increase in absolute magnitude as we incorporate later retirement ages. When we include retirement ages 60, 61, and 62—ages at which the 1951 cohort can retire but the 1952 cohort cannot—we observe a substantial increase in standard errors. For the word recall measures, statistical significance begins to emerge when we include women who retire at age 63, the ERA for the 1952 cohort. As we continue adding older retirement ages, the coefficients remain significant and become more robust.

These findings suggest that the significant negative effects observed in Table 2 are not driven by the potentially misspecified 1951 cohort. Instead, they likely reflect the true effect revealed by the correctly specified 1952 cohort.

Figure 2: Analysis of 1952 cohort

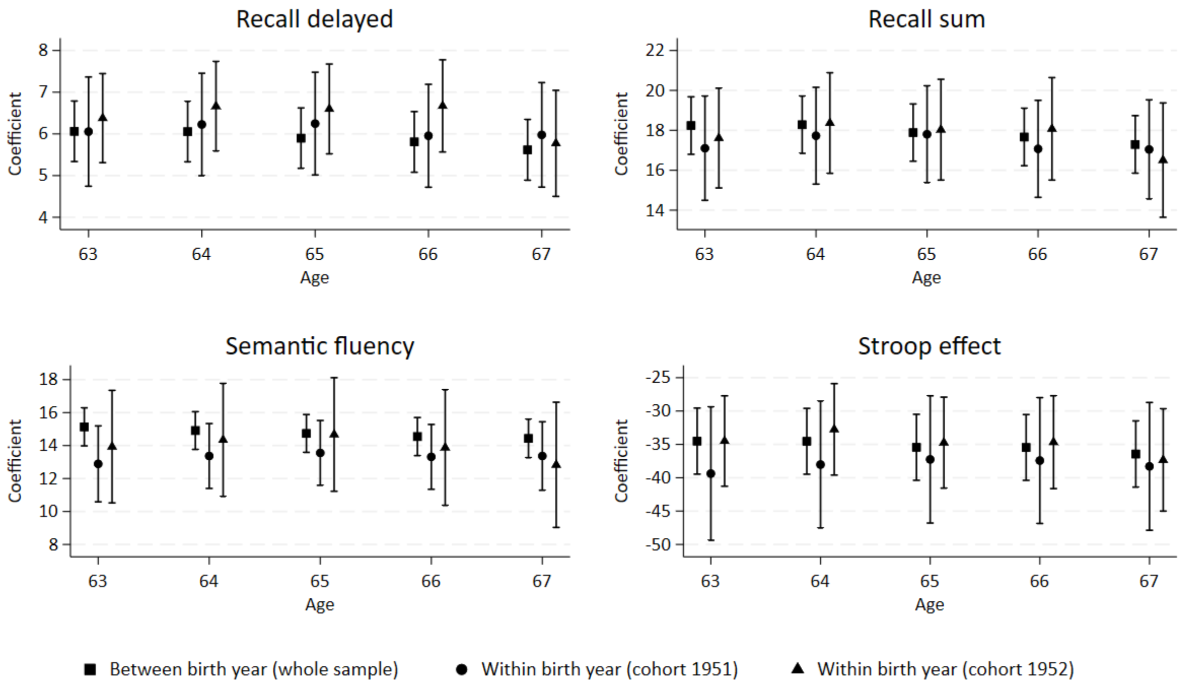


Notes: This figure shows calculations of Equation 1 with full controls and with different samples regarding the 1952 cohort. We always include women not yet retired. We gradually increase the sample size with an increasing retirement age. Point 53 indicates that we drop the entire 1952 cohort. Point 67 includes the whole 1952 cohort and gives results from Table 2. Estimates are shown including 95% confidence intervals. Appendix Figure A8 shows placebo estimations of the same analysis when dropping the 1953 cohort.

4.4.2 Cohort heterogeneity

To demonstrate that age effects within our estimation are not driven by potential cohort effects, we follow Salthouse (2013) who shows that within-cohort differences in cognitive abilities across ages are as large as between-cohort differences across ages. This finding was obtained using a similarly structured dataset. In a cross-section, observed at different time periods, a proper separation of age effects and cohort effects is not possible. Therefore, the author compares age effects by evaluating birth year effects while holding the test year constant (between-cohort differences) and age effects by evaluating test year effects while holding birth year constant (within-cohort differences). In Figure 3 we adapt this analysis to our empirical set-up regarding the NAKO-analysis. We graphically plot coefficient estimates of age dummies that are computed with the whole set of birth years (between-cohort differences) and age dummies that are computed within a single birth cohort (within-cohort differences). As we observe the widest range of age for birth cohorts 1951 and 1952, that is ages 63 to 67, we focus only on these birth cohorts to show within-cohort differences. We separately regress the cognition outcomes on age dummies of both samples without including a constant and plot coefficients and 95% confidence intervals. We also include full controls. Estimates in Figure 3 show that for each cognition outcome age effects of within- and between-cohort differences are predominantly not statistically different from each other. We therefore assume that controlling for age dummies that simultaneously capture cohort effects does not bias our results.

Figure 3: Within- and between-cohort differences across age



Notes: This figure shows age coefficients calculated following Salthouse (2013). We regress the cognition measures on age dummies separately with three specifications: within the whole sample, only for birth cohort 1951 and only for birth cohort 1952. We control for birthplace Germany, education dummies, and study center FE. Estimates are shown including 95% confidence intervals.

Similarly, we examine our results of column 5 in Table 2 to determine whether possible trends between the birth cohorts may play a role for our estimation. Figure A7 shows coefficients for each age year included in our sample from three separate estimations. Identical to Figure 3 we do not find any significant differences in the risk of a diagnosis between the full sample, the 1951 cohort and the 1952 cohort at any age. We therefore assume that controlling for age fixed effects that simultaneously capture cohort effects does not bias our results.

4.4.3 Wave effects

In our specifications, observations from the 2014 wave are only available for birth years not affected by the reform, specifically ages 63 (born in 1951) to 65 (born in 1949). Conversely, observations from the 2019 wave exclusively capture birth years affected by the reform, namely ages 65 (born in 1954) to 67 (born in 1952). This discrepancy raises the concern of potential bias stemming from wave-specific effects in 2014 and 2019, which could confound our estimates. To address this potential source of bias, we exclude both the 2014 and 2019 waves and re-estimate the models to examine whether the results change. The findings, presented in Table 4, show that the OLS estimates, as well as the first- and second-stage estimates, remain consistent with those in Table 2, with no significant changes. The second-stage estimates suggest that the effects on word recall and the Stroop effect remain statistically significant, while the effects on semantic fluency are no longer significant.

Table 4: Excluding waves 2014 and 2019

	1949-1951 vs. 1952-1954				
	Delayed word recall (1)	Sum of words recalled (2)	Semantic fluency (3)	Stroop effect (4)	Diagnosis of CD (5)
OLS					
Years retired	-0.021**	-0.054**	-0.039	-0.070	0.00017***
	(0.011)	(0.022)	(0.028)	(0.061)	0.00005
First stage					
Birth cohort $\geq$ 1952	-0.586***	-0.585***	-0.577***	-0.583***	-0.862***
	(0.074)	(0.074)	(0.074)	(0.074)	0.006
2SLS					
Years retired	-0.287**	-0.526**	-0.357	-1.396*	-0.00013
	(0.124)	(0.257)	(0.326)	(0.738)	(0.00036)
Controls	Yes	Yes	Yes	Yes	No
Study center FE	Yes	Yes	Yes	Yes	No
Age FE	Yes	Yes	Yes	Yes	Yes
Observations	8,189	8,185	8,276	8,185	1,443,280

Note: This table shows coefficients for the outcome variables of cognitive abilities and cognitive disease. Compared to Table 2 individuals from waves 2014 and 2019 are excluded here as a robustness check for potential wave effects. The first row shows the estimated coefficients from an OLS regression based on Equation 1. The second row shows the results from the first stage and the third row from the corresponding 2SLS estimation. The regressions in column 1 to 4 include a dummy for birthplace Germany and education dummies according to the ISCED standards. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5 Conclusion

In this study, we investigate the impact of retirement duration on women's cognitive abilities and the diagnosis of cognitive decline (CD). We use data from the NAKO study, supplemented by health insurance claims data, and exploit the 1999 German pension reform, which raised the early retirement age (ERA) for women by three years, as an instrument to measure these effects. This approach is crucial because retirement decisions are often influenced by personal choice and may be subject to reverse causality, particularly when declining mental health influences the decision to retire.

Our findings indicate that an additional year in retirement has a significant negative effect on the cognitive abilities of the group of compliers—women whose retirement decisions were altered by the pension reform. This effect is specifically observed in cognition measures related to fluid intelligence. However, up to the age of 67, we find no significant effect of retirement duration on the risk of being diagnosed with a CD, which includes conditions such as dementia and mild cognitive impairment (MCI). It is possible that such effects may emerge at older ages, suggesting that the relationship between retirement and cognitive decline may become more pronounced later in life..

The findings of our study contribute to the literature by utilizing data sources that allow for a focused analysis of a single country, Germany. Effects from instruments utilizing different pension system from various countries can therefore be neglected. Moreover, the instrument we use defines a direct exogenous shock to the decision to retire of the women in our sample. To the best of our knowledge, we are the first to study causal effects on cognitive diseases within a specific age group and find null effects. A limitation is that we cannot observe any long-term effects, as women born around the 1952 cohort are not yet expected to suffer from severe cognitive diseases. In the future, we tend to investigate whether the duration of retirement affects the diagnosis of cognitive diseases in the relevant birth cohorts beyond our defined age range. However, based on our results, we conclude that an extension of the retirement entry to a later date can potentially mitigate the decline of cognitive abilities induced by additional years in retirement. In the short term, it does not reduce the risk of developing cognitive diseases.

References

American Psychological Association (1995). Intelligence: Knowns and Unknowns, Report of a task force convened by the American Psychological Association. *Science Directorate, Washington DC [LGH]*.

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

Bai, W., Chen, P., Cai, H., Zhang, Q., Su, Z., Cheung, T., Jackson, T., Sha, S., and Xiang, Y.-T. (2022). Worldwide prevalence of mild cognitive impairment among community dwellers aged 50 years and older: a meta-analysis and systematic review of epidemiology studies. *Age and Ageing*, 51(8):afac173.

Banks, J. and Mazzonna, F. (2012). The Effect of Education on Old Age Cognitive Abilities: Evidence from a Regression Discontinuity Design. *The Economic Journal*, 122(560):418–448.

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.

Ben-David, B. M., Tewari, A., Shakuf, V., and Van Lieshout, P. H. (2014). Stroop effects in Alzheimer's disease: selective attention speed of processing, or color-naming? A meta-analysis. *Journal of Alzheimer's Disease*, 38(4):923–38.

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

Case, A. and Paxson, C. (2010). Causes and consequences of early-life health. *Demography*, 47:65 – 85.

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

Christelis, D., Jappelli, T., and Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review*, 54(1):18–38.

Coe, N. B., von Gaudecker, H. M., Lindeboom, M., and Maurer, J. (2012). The effect of retirement on cognitive functioning. *Health Economics*, 21(8):913–27.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, 29(1):1–28.

DGN e. V. and DGPPN e. V. (2023). S3-Leitlinie Demenzen.

Doblhammer, G., Fink, A., and Fritze, T. (2015). Short-term trends in dementia prevalence in Germany between the years 2007 and 2009. 11(3):291–299.

Dufouil, C., Pereira, E., Chêne, G., Glymour, M. M., Alpérovitch, A., Saubusse, E., Risse-Fleury, M., Heuls, B., Salord, J.-C., Brieu, M.-A., and Forette, F. (2014). Older age at retirement is associated with decreased risk of dementia. *European Journal of Epidemiology*, 29(5):353–361.

Etgeton, S., Fischer, B., and Ye, H. (2023). The Effect of Increasing Retirement Age on Households' Savings and Consumption Expenditure. *Journal of Public Economics*, 221.

Federal Statistical Office (2024). Erwerbstätigkeit von Frauen: Deutschland mit dritthöchster Quote in der EU. [Online] URL: https://www.destatis.de/Europa/DE/Thema/Bevoelkerung-Arbeit-Soziales/Arbeitsmarkt/ArbeitsmarktFrauen.html. accessed 23. Oktober 2024.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: what IQ tests really measure. *Psychological bulletin*, 101(2):171.

Freise, D., Schmitz, H., and Westphal, M. (2022). Late-career unemployment and cognitive abilities. *Journal of Health Economics*, 86:102689.

Fürstenau, E., Gohl, N., Haan, P., and Weinhardt, F. (2023). Working life and human capital investment: Causal evidence from a pension reform. *Labour Economics*, 84.

Georges, D., Rakusa, E., Holtz, A.-V., Fink, A., and Doblhammer, G. (2023). Dementia in Germany: epidemiology, trends and challenges. 8(3):30–48. Place: Germany.

Gerstorf, D., Ram, N., Hoppmann, C., Willis, S. L., and Schaie, K. W. (2011). Cohort differences in cognitive aging and terminal decline in the Seattle Longitudinal Study. *Developmental psychology*, 47(4):1026 – 1041.

Geyer, J., Barschkett, M., Haan, P., and Hammerschmid, A. (2023). The effects of an increase in the retirement age on health care costs: evidence from administrative data. *The European Journal of Health Economics*, 24(7):1101–1120.

Geyer, J., Haan, P., Hammerschmid, A., and Peters, M. (2020). Labor Market and Distributional Effects of an Increase in the Retirement Age. *Labour Economics*, 65.

Geyer, J. and Welteke, C. (2021). Closing Routes to Retirement for Women. *Journal of Human Resources*, 56(1):311–341.

Grobe, T. G. and Szecsenyi, J. (2023). *Kindergesundheit: Frühgeburtlichkeit und Folgen, series = Barmer Arztreport*. Barmer Institut für Gesundheitssystemforschung.

Grotz, C., Letenneur, L., Bonsang, E., Amieva, H., Meillon, C., Quertemont, E., Salmon, E., Adam, S., and ICTUS/DSA group (2015). Retirement Age and the Age of Onset of Alzheimer's Disease: Results from the ICTUS Study. *PLOS ONE*, 10(2):e0115056.

Hagen, J. (2017). The effects of increasing the normal retirement age on health care utilization and mortality. *Journal of Population Economics*, 31(1):193–234.

Horner, E. M. and Cullen, M. R. (2016). The impact of retirement on health: quasi-experimental methods using administrative data. *BMC Health Services Research*, 16(1).

Karlsson, P., Thorvaldsson, V., Skoog, I., Hudmundsson, P., and Johansson, B. (2015). Birth cohort differences in fluid cognition in old age: Comparisons of trends in levels and change trajectories over 30 years in three population-based samples. *Psychology and Aging*, 30(1):89 – 94.

Kasteridis, P., Mason, A. R., Goddard, M. K., Jacobs, R., Santos, R., and McGonigal, G. (2015). The Influence of Primary Care Quality on Hospital Admissions for People with Dementia in England: A Regression Analysis. 10(3):e0121506.

Kleinedamm, L., Stark, M., Riedel-Heller, S. G., Pabst, A., Schmiedek, F., Streit, F., Rietschel, M., Klinger-König, J., Grabe, H. J., Erhardt, A., Gelbrich, G., Schmidt, B., Investigators, N., Berger, K., and Wagner, M. (2022). The assessment of cognitive function in the German

National Cohort (NAKO)–Associations of demographics and psychiatric symptoms with cognitive test performance. *The World Journal of Biological Psychiatry*, pages 1–15.

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.

Li, Y., Baldassi, M., Johnson, E. J., and Weber, E. U. (2013). Complementary cognitive capabilities, economic decision making, and aging. *Psychology and aging*, 28(3):595.

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

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

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., and Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, 38(1):115–142.

McFadden, D. (2008). Human Capital Accumulation and Depreciation. *Applied Economic Perspectives and Policy*, 30(3):379–385.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive psychology*, 41(1):49–100.

Mosca, I. and Wright, R. E. (2018). Effect of Retirement on Cognition: Evidence From the Irish Marriage Bar. *Demography*, 55(4):1317–1341.

Mueller, K. D., Koscik, R. L., LaRue, A., Clark, L. R., Hermann, B., Johnson, S. C., and Sager, M. A. (2015). Verbal Fluency and Early Memory Decline: Results from the Wisconsin Registry for Alzheimer's Prevention. *Archives of Clinical Neuropsychology*, 30(5):448–57.

Murphy, K. J., Rich, J. B., and Troyer, A. K. (2006). Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of Alzheimer's type dementia. *Journal of the International Neuropsychological Society*, 12(4):570–4.

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

Nikolai, T., Bezdicek, O., Markova, H., Stepankova, H., Michalec, J., Kopecek, M., Dokoupilova, M., Hort, J., and Vyhnalek, M. (2018). Semantic verbal fluency impairment is detectable in patients with subjective cognitive decline. *Applied Neuropsychology: Adult*, 25(5):448–457.

OECD (2018). *Dementia prevalence*. OECD Publishing.

Peters, A., German National Cohort, C., Peters, A., Greiser, K. H., Gottlicher, S., Ahrens, W., Albrecht, M., Bamberg, F., Barnighausen, T., Becher, H., Berger, K., Beule, A., Boeing, H., Bohn, B., Bohnert, K., Braun, B., Brenner, H., Bulow, R., Castell, S., Damms-Machado, A., Dorr, M., Ebert, N., Ecker, M., Emmel, C., Fischer, B., Franzke, C. W., Gastell, S., Giani, G., Gunther, M., Gunther, K., Gunther, K. P., Haerting, J., Haug, U., Heid, I. M., Heier, M., Heinemeyer, D., Hendel, T., Herbolsheimer, F., Hirsch, J., Hoffmann, W., Holleczek, B., Holling, H., Horlein, A., Jockel, K. H., Kaaks, R., Karch, A., Karrasch, S., Kartschmit, N., Kauczor, H. U., Keil, T., Kemmling, Y., Klee, B., Kluppelholz, B., Kluttig, A., Kofink, L., Kottgen, A., Kraft, D., Krause, G., Kretz, L., Krist, L., Kuhnisch, J., Kuss, O., Legath, N., Lehnich, A. T., Leitzmann, M., Lieb, W., Linseisen, J., Loeffler, M., Macdonald, A., Maier-Hein, K. H., Mangold, N., Meinke-Franze, C., Meisinger, C., Melzer, J., Mergarten, B., Michels, K. B., Mikolajczyk, R., Moebus, S., Mueller, U., Nauck, M., Niendorf, T., Nikolaou, K., Obi, N., Ostrzinski, S., Panreck, L., Pigeot, I., Pischon, T., Pschibul-Thamm, I., Rathmann, W., Reineke, A., Roloff, S., Rujescu, D., Rupf, S., Sander, O., Schikowski, T., Schipf, S., Schirmacher, P., Schlett, C. L., Schmidt, B., Schmidt, G., et al. (2022). Framework and baseline examination of the German National Cohort (NAKO). *European Journal of Epidemiology*, 37(10):1107–1124.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., and Kokmen, E. (1999). Mild Cognitive Impairment: Clinical Characterization and Outcome. *Archives of Neurology*, 56(3):303.

Rechel, B., Grundy, E., Robine, J. M., Cylus, J., Mackenbach, J. P., Knai, C., and McKee, M. (2013). Ageing in the European Union. *The Lancet*, 381(9874):1312–22.

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

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, 30(4):507–14.

Salthouse, T. A. (2013). Within-cohort age-related differences in cognitive functioning. *Psychological Science*, 24(2):123–30.

Schiele, V. and Schmitz, H. (2023). Understanding cognitive decline in older ages: The role of health shocks. *European Economic Review*, 151.

Schmitz, H. and Westphal, M. (2021). The Dynamic and Heterogeneous Effects of Retirement on Cognitive Decline. *Ruhr Economic Papers*, 918.

Schmitz, H. and Westphal, M. (2022). Early and later-life stimulation: How retirement shapes the effect of education on old-age cognitive abilities. *Not yet published working paper. Current status: under review*.

Schneeweis, N., Skirbekk, V., and Winter-Ebmer, R. (2014). Does education improve cognitive performance four decades after school completion? *Demography*, 51(2):619–43.

Spieler, D. H., Balota, D. A., and Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human perception and performance*, 22(2):461.

Sundström, A., Rönnlund, M., and Josefsson, M. (2020). A nationwide Swedish study of age at retirement and dementia risk. *International Journal of Geriatric Psychiatry*, 35(10):1243–1249.

Tymula, A., Rosenberg Belmaker, L. A., Ruderman, L., Glimcher, P. W., and Levy, I. (2013). Like cognitive function, decision making across the life span shows profound age-related changes. *Proceedings of the National Academy of Sciences*, 110(42):17143–8.

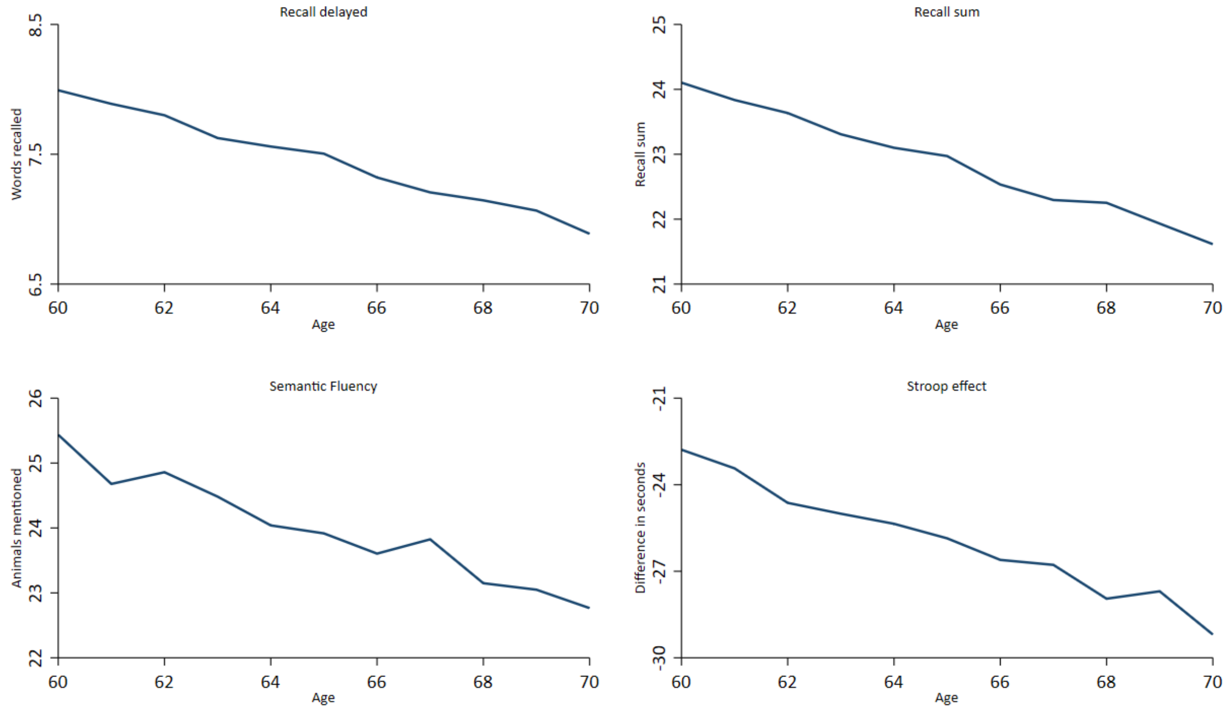
West, R. and Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, 37(2):179–189.

Zelinski, E. M. and Kennison, R. F. (2007). Not Your Parents' Test Scores: Cohort Reduces Psychometric Aging Effects. *Psychology and Aging*, 22(3):546 – 557.

Zentralinstitut kassenärztliche Versorgung (2024). Demenz und andere Kognitive Defizite: Zi-Kodier-Manual.

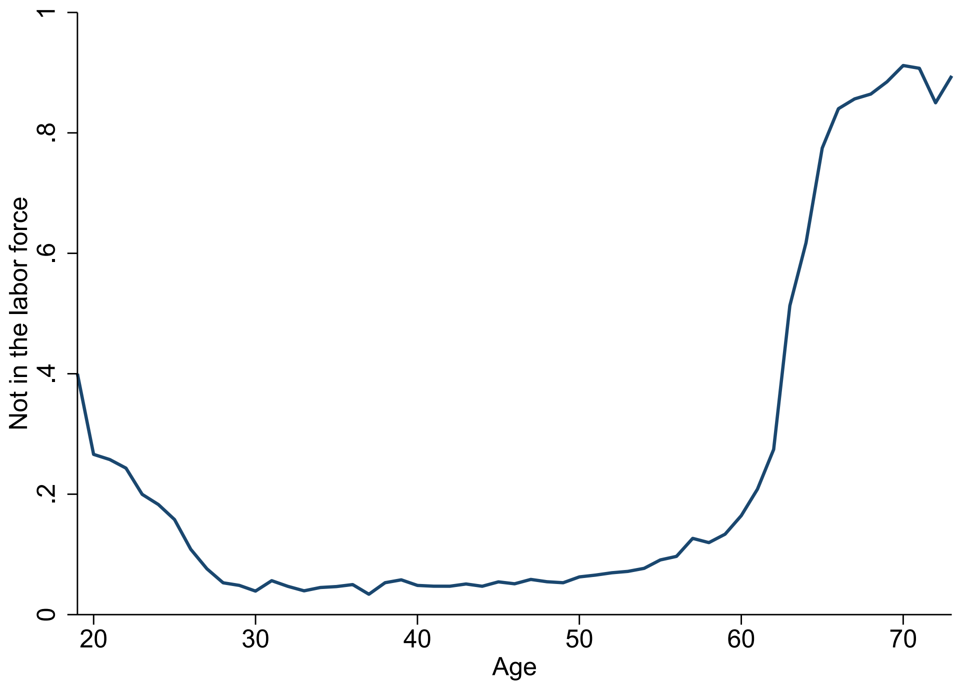
Appendix: Additional tables and figures

Figure A1: Cognitive abilities of women by age



Notes: This table shows cognitive abilities of women in terms of test scores by age in the NAKO dataset.

Figure A2: Labor force participation among females by age



Notes: This figure shows the share of women not participating in the labor force by age in the NAKO dataset.

Table A1: ICD - Codes

Code	**Description**
Dementia	
G30	Alzheimer's disease
G31.0	Frontotemporal dementia
G31.82	Lewy body disease
G23.1	Progressive supranuclear ophthalmoplegia (Steele-Richardson-Olszewski syndrome)
F00	Dementia in Alzheimer's disease
F01	Vascular dementia
F02	Dementia in other diseases classified elsewhere
F03	Unspecified dementia
F05.1	Delirium superimposed on dementia
Mild Cognitive Impairment	
F06.7	Mild Cognitive Disorder

Notes: This table shows the ICD-10-GM codes used for the outcome variable. ICD selection as in Kasteridis et al. (2015); Doblhammer et al. (2015); Georges et al. (2023); Zentralinstitut kassenärztliche Versorgung (2024); DGN e. V. and DGPPN e. V. (2023).

Table A2: Summary statistics by reform eligibility of NAKO dataset

	Born 1949-1951 ($n = 5,120$)		Born 1952-1954 ($n = 4,718$)	
	Mean	St. dev.	Mean	St. dev.
Socioeconomic factors				
Age	65.72	1.06	64.05	1.00
Birth year	1950.20	0.79	1952.76	0.77
Retired (in %)	86.81	33.85	65.58	47.51
Retirement age	62.21	2.36	62.09	2.19
Years retired	4.02	2.73	2.08	2.30
Born in Germany (in %)	90.95	28.69	89.96	30.05
Primary education (in %)	0.86	9.22	0.48	6.90
Lower secondary education (in %)	3.76	19.03	3.92	19.42
Upper secondary education (in %)	30.00	45.83	30.95	46.23
University entrance qualification (in %)	19.21	39.39	18.02	38.44
University degree (in %)	43.69	49.61	44.20	49.67
Doctorate (in %)	2.49	15.58	2.44	15.43
Cognitive abilities				
Delayed word recall	7.36	2.22	7.61	2.22
Word recall Sum	22.66	4.77	23.23	4.77
Semantic fluency	23.86	6.28	24.23	6.41
Stroop effect	26.35	14.25	24.96	12.36
Observations				
Delayed Word Recall	4,835		4,458	
Word Recall Sum	4,834		4,455	
Stroop effect	4,834		4,452	

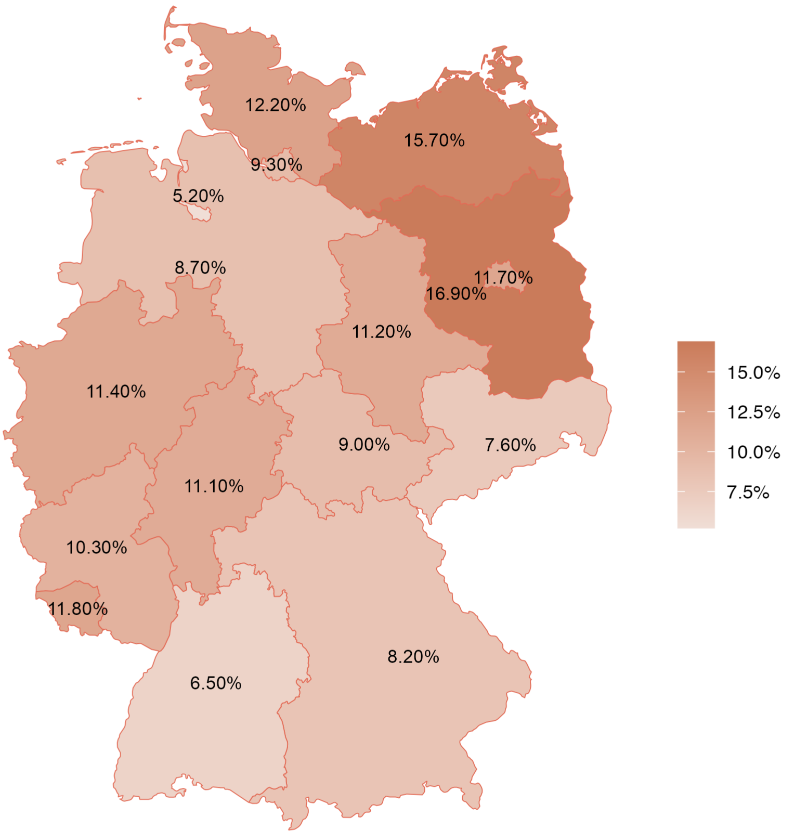
Note: This table shows descriptive statistics for individuals in the NAKO dataset.

Table A3: Summary statistics by reform eligibility of administrative dataset

	Born 1949-1951 ($n = 1,053,495$)		Born 1952-1954 ($n = 799,609$)	
	Mean	St. dev.	Mean	St. dev.
Socioeconomic factors				
Age	64.98	1.41	64.57	1.26
Retired (in %)	0.73	0.45	0.58	0.49
Birth year	1949.99	0.81	1952.83	0.80
Retirement age	63.19	2.17	64.14	1.42
Years retired	2.54	2.27	1.32	1.45
Health				
Dementia or MCI	0.0066	0.081	0.0059	0.076
Yearly Hospital Costs	791.99	4142.25	845.52	4493.79
Elixhauser Score				
< 0	0.38	0.49	0.41	0.49
$= 0$	0.25	0.44	0.21	0.41
$1 - 4$	0.09	0.29	0.10	0.29
> 4	0.28	0.45	0.29	0.45

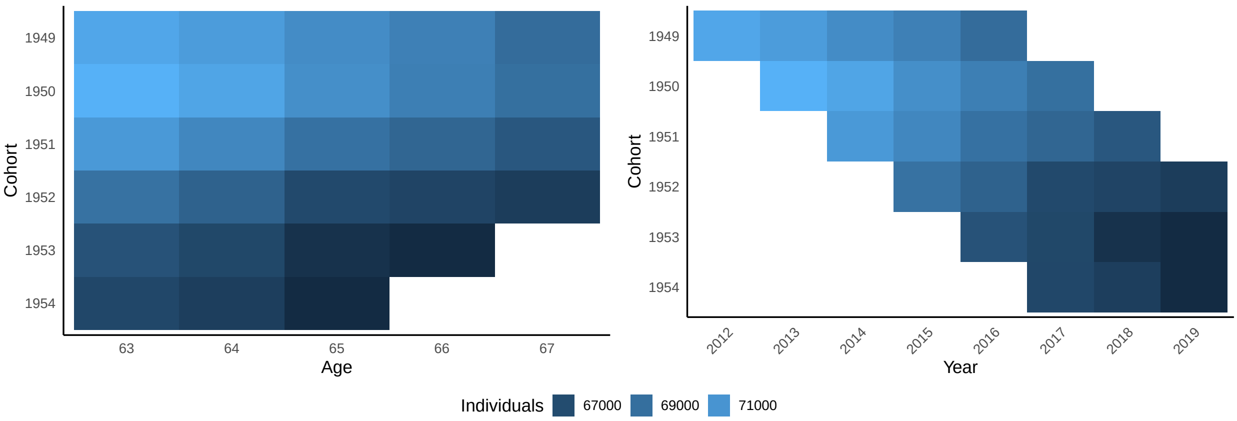
Note: This table displays descriptive statistics for individuals born between 1949 to 1954 in the years 2012 to 2019 from the administrative dataset.

Figure A3: Regional distribution of insured individuals



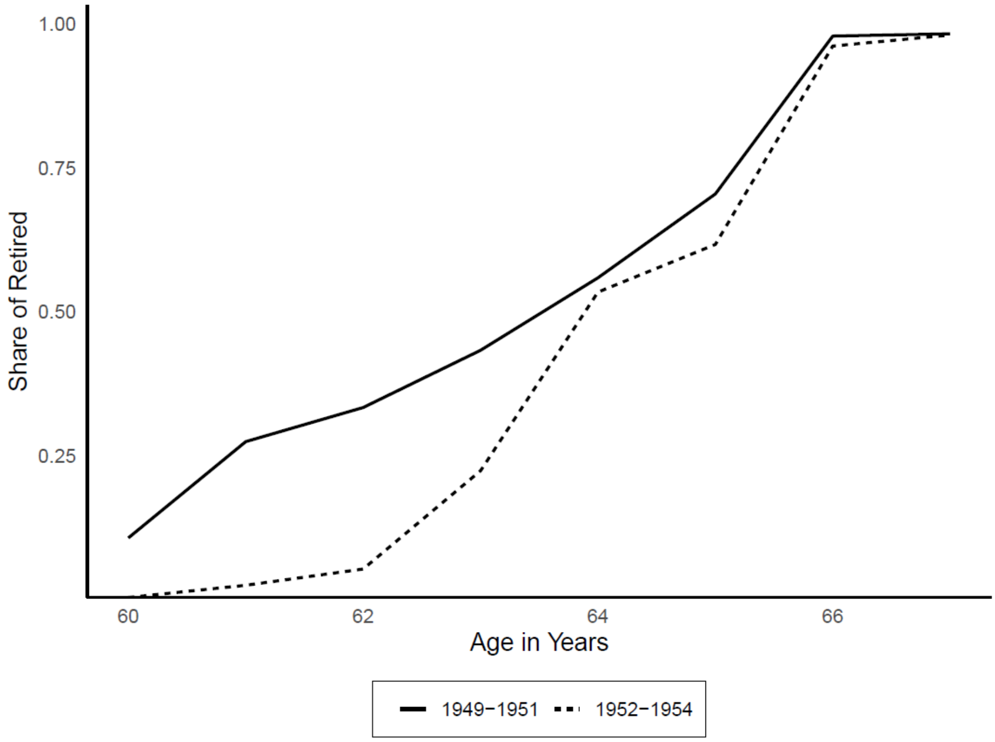
Note: This figure displays the share of individuals in each state in the administrative dataset (Grobe and Szecsenyi, 2023).

Figure A4: Sample distribution across age, cohort and year



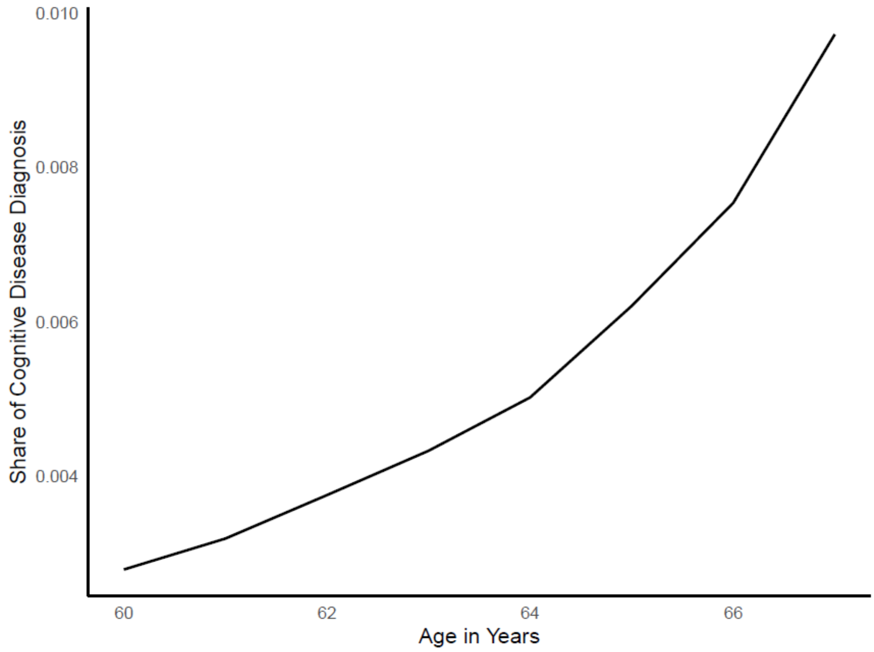
Note: This figure shows the number of observations by age and year of birth in the left graph and by year and year of birth in the right graph.

Figure A5: Retirement status by age and cohort in administrative data



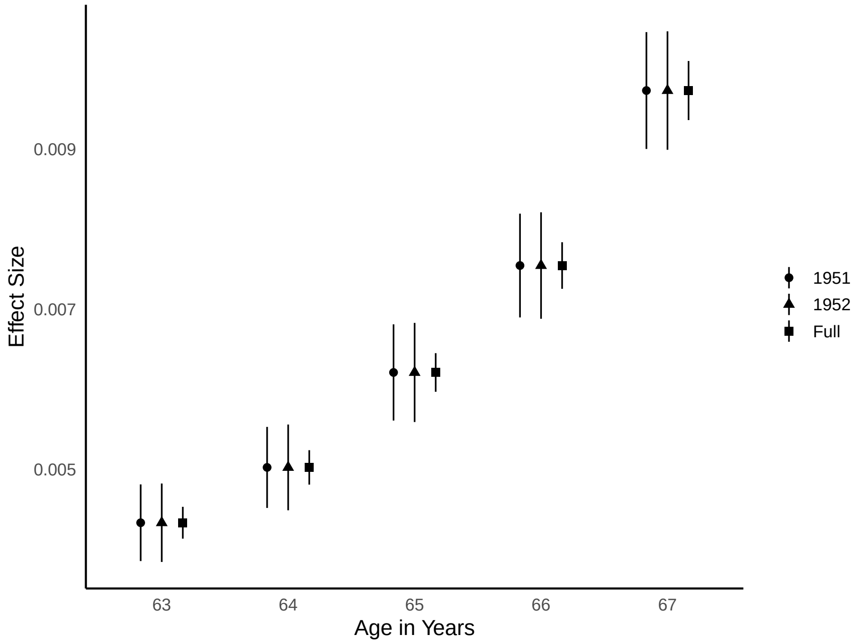
Note: This figure displays the share of retired women by age and year of birth.

Figure A6: Dementia or MCI by age



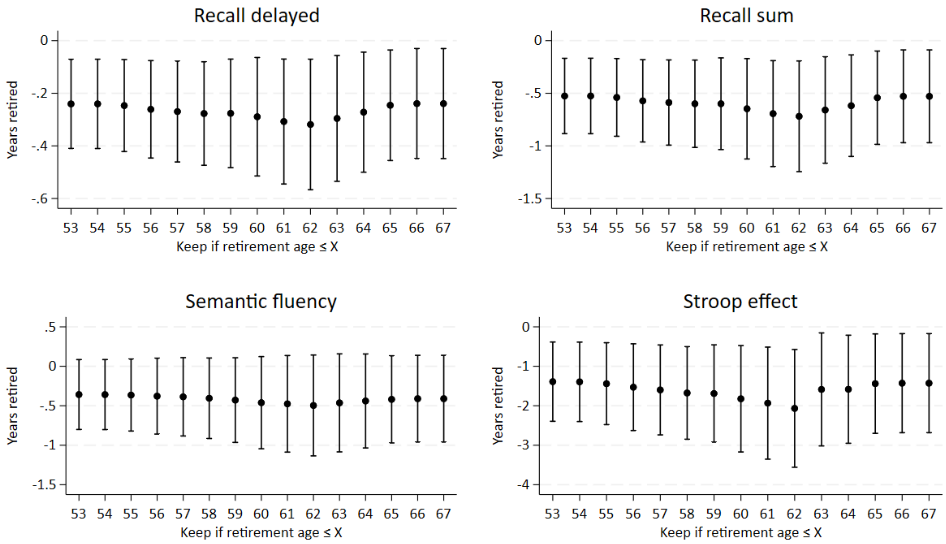
Note: This figures shows the probability of a MCI or dementia diagnosis by age. Diagnoses are defined as described in Table. Diagnoses are defined as shown in Table A1.

Figure A7: Cohort heterogeneity



Note: This figure shows a comparison of our full sample to birth cohorts 1951 and 1952 in terms of risk of a dementia or MCI diagnosis following Salthouse (2013). We regress the administrative diagnoses on age dummies separately with three specifications: within the whole sample, only for birth cohort 1951 and only for birth cohort 1952. Estimates are shown including 95% confidence intervals.

Figure A8: Placeob test - analysis of 1953 cohort



Notes: This figure shows calculations of Equation 1 with full controls and with different samples regarding the 1953 cohort. We always include women not yet retired. We gradually increase the sample size with an increasing retirement age. Point 53 indicates that we drop the entire 1953 cohort. Point 67 includes the whole 1953 cohort. Estimates are shown including 95% confidence intervals.