Effects of Retirement on Cognitive Functioning – Evidence from Biomedical and Administrative Insurance Claims Data*

Henrik Bergschneider
RWI
Paderborn University

Robin Kottmann
RWI
Paderborn University

Hendrik Schmitz
RWI
Paderborn University

Matthias Westphal
University of Hagen
RWI

Abstract

We study the effects of retirement on cognitive functioning among women aged 63 to 67 by exploiting a German retirement reform that raised the early retirement age for women born after 1951 by three years, from 60 to 63. Our indicators of cognitive functioning are experimental measures (word recall, semantic fluency, and the Stroop test) from a large biomedical data set, as well as the diagnosis of cognitive disorders from administrative health insurance claims. We find reductions of around 12% of a standard deviation per year in retirement for measures of fluid intelligence and of an insignificant 6% for crystallized intelligence. The diagnosis of cognitive disorders remains unaffected.

Keywords: Cognitive abilities, Retirement, Pension Reform
JEL Classification: *C31, J14, J24*

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

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

Hendrik Schmitz: RWI – Leibniz-Institut für Wirtschaftsforschung, Hohenzollernstr. 1-3, 45128 Essen, Germany, Tel.: +49 201 / 8149-301, E-mail: hendrik.schmitz@rwi-essen.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 Gunther Bensch, Peter Haan and the participants of the Annual Meeting of the German Health Economics Association (DGGÖ) in Halle, the EUHEA PhD conference in Lucerne, the Workshop of the DGGÖ Health Econometrics Working Group in Berlin, 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 individual decision-making across the lifespan (Tymula et al., 2013; Li et al., 2013) involving, for instance, financial decisions or health behavior (Christelis et al., 2010; Cutler and Lleras-Muney, 2010). However, it is widely known that cognitive abilities decline with age (Salthouse, 2009), while the prevalence of conditions such as mild cognitive impairment (MCI) or dementia increases. The ongoing demographic change will lead to a rise in the population with cognitive impairment, putting even more pressure on already stressed health and long-term care systems (Rechel et al., 2013). Many studies from various disciplines examine whether this age-related decline is inevitable or malleable, ideally to derive policy recommendations to address it.

We investigate how retirement – a major choice in an individual's life – adds to cognitive decline. Rohwedder and Willis (2010) argue that retirees face fewer mental challenges than labor force participants, particularly when they leave a cognitively stimulating work environment. This is known as the "use-it-or-lose-it hypothesis". Our study is not the first to examine this. Dufouil et al. (2014), Grotz et al. (2015) and Sundström et al. (2020) find a correlation between retirement age and risk of dementia. Individuals who retire later (and spend less time in retirement) are less likely to receive a dementia diagnosis. In terms of causal claims, the recent economic literature has focused particularly on the effects of retirement on measures of cognition, such as the word recall test. Rohwedder and Willis (2010) and Bonsang et al. (2012) suggest that retirement negatively affects cognitive abilities. Other studies confirm these findings and additionally find that length of retirement increases this negative effect (Celidoni et al., 2017; Mazzonna and Peracchi, 2012, 2017). Furthermore, there is heterogeneity as regards the type of former work (Mazzonna and Peracchi, 2017; Coe et al., 2012) and also for gender (Atalay et al., 2019; Mazzonna and Peracchi, 2012; Mosca and Wright, 2018). Schmitz and Westphal (2021) find heterogeneous negative average effects of retirement on cognitive abilities 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.

We believe that we can contribute substantially to the existing literature for two reasons. First, most of the established literature in this field relies on cross-country differences in retirement regulations for identification (exceptions are Atalay et al., 2019, for Australia and Mosca and Wright, 2018, for Ireland). This approach has been criticized for potentially conflating retirement effects with other cross-country differences that may influence cognition. In contrast, we exploit a German pension reform of 1999 that introduced a sharp increase in early retirement age (ERA) for women from 60 (until birth cohort 1951) to 63 (as of birth cohort 1952). This reform led to a discontinuous and substantial decline in retirement rates of affected women, as shown by Geyer and Welteke (2021). Hence, we add to the literature by imposing a different set of assumptions than most previous work in studying the effects of retirement on cognition. Second, while all previous literature lies exclusively on survey data, we are, to the best of our knowledge, the first to use administrative data and diagnoses of cognitive disease as an outcome in a study that exploits exogenous variation in estimating retirement effects on cognition.[1]

We use data from two different sources. The first is the German National Cohort (NAKO), an ongoing examination that, in its first wave (2014–2019), carried out medical evaluations of 205,000 individuals aged between 20 and 69. We use the NAKO to estimate retirement effects on neuropsychological cognition measures. These are word recall, semantic fluency, and the Stroop effect. Word recall and Stroop effect are considered measures of fluid intelligence, that is, very briefly, the ability to solve novel reasoning problems, abstract and logical thinking. Semantic fluency also includes learned procedures and knowledge and, thus, also reflects crystallized intelligence. The second source is a large, nationally representative dataset from a German health insurer covering about 10% of the German population from 2012–2019. We utilize this information to explore diagnoses of cognitive disorders (CD) such as dementia and MCI. Our analysis is based on a sample of German women born 1949–1954, that is, around the pivotal cohort affected by the retirement reform. Cognition is measured at ages 63–67. Each data source has its advantages and limitations, and we argue that combining them offers a comprehensive view of the medium-term effects of retirement on cognition. The administrative dataset provides a large sample size (more than 400,000 women in the mentioned birth cohorts) and detailed information on diagnoses and retirement behavior. However, dementia, while not completely uncommon among individuals in their 60s usually develops – and, possibly more important, is diagnosed – at later ages (OECD, 2018). Here, the experimentally retrieved measures of cognition from the NAKO are crucial, as they allow to observe medium-run effects on cognition that are below the level of clinical diagnoses. Yet, the sample size for women born between 1949 and 1954 is smaller (around 9,000), and retirement status is self-reported in that data set.

We find that an additional year spent in retirement reduces measures of fluid intelligence (word recall and Stroop test results) significantly by around 12% of a standard deviation among the studied women. The measure of crystallized intelligence (semantic fluency) is reduced by around 6 percent of a standard deviation, which is not statistically significant. We do not find significant effects on medical outcomes related to cognitive health. The probability of receiving a relevant diagnosis of a cognitive disorder increases by just 0.011 percentage points per year of retirement, compared to an overall prevalence of 0.6 percent within the studied age range. However, our findings do not preclude the possibility that the effects on cognitive disorders may become more pronounced later in life. All

[1]More generally, our paper also relates to the literature on the impact of old-age retirement on health. Here, some studies utilize diagnoses and drug prescriptions from public health insurance administrative data 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).

in all, our results of retirement effects on the cognition of women are in line with some studies (Mazzonna and Peracchi, 2012, 2017; Schmitz and Westphal, 2021) that also find effects but diverge from others (Mosca and Wright, 2018; Atalay et al., 2019) that do not find such effects. It is important to note that our results identify local average treatment effects for a specific group of women within a defined age range. We do not claim that these effects are generalizable to other cohorts, different types of compliers, or across countries. Nevertheless, our results imply that postponing retirement ages, which is done and discussed in almost all countries to sustain financial stability of pay-as-you-go retirement schemes – and which is exceptionally unpopular – might also bring positive aspects in some dimensions for individuals.

The paper is structured as follows. Section 2 describes the background of the pension reform for women that we use as exogenous variation. Section 3 presents the data and the variables used. The estimation strategy is outlined in Section 4, followed by the results in Section 5. Section 6 concludes.

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 allows to retire early. The German pension 1999 reform 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 age 60 while the ORA was 65.[2] This regulation was commonly known as the "pension for women." The precondition to enter early retirement included coverage by the statutory pension insurance for at least 15 years and contributions to the pension fund for 10 years after their 40th birthday.[3] Women who choose to retire early face a permanent pension benefit deduction of 0.3 % for each month they undercut the ORA. Before the pension reform came into force, women who retired at 60 faced a permanent deduction from their pension benefits, which accumulated to 18%.

The reform of 1999 abolished the pension for women, intending to increase female 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, could still retire at 60. Consequently, this reform implied a strict demarcation regarding the ERA for different

[2] By opportunity to retire early we mean the opportunity to retire due to age with the receipt of retirement benefits. Of course, individuals can retire at any time without retirement benefits. Moreover, retirement due to disability is possible at younger ages, too.

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

birth cohorts. Geyer and Welteke (2021) examine the effects of the reform on different labor market outcomes. They find that the pension reform increased employment rates by 13.5 percentage points among women aged 60–62 who were affected by the reform. This is a 30% increase compared to the pre-reform mean. Geyer et al. (2020) also investigate the labor market effects of the reform. They find that the reform led to a significant drop in retirement rates for the affected birth cohorts. 60-62-year-old women born after the cutoff are 16 percentage points less likely to be retired, which equals a 50% drop compared to pre-reform means. Moreover, they report an increase in employment rates by 8.4 percentage points. Hence, women born after January 1, 1952, extended their labor force participation compared to women born before. We build on this by instrumenting the years spent in retirement by the reform.

Other studies analyze the effects of the German pension reform on various further outcomes. Etgeton et al. (2023) find that households decrease savings rates and increase leisure spending in response to the reform. Households with married women drive these effects. Fürstenau et al. (2023) show the positive effect of the reform on participation in on-the-job training with a sharp RD design indicating positive employment effects. The pension reform also has been used to show effects on health (care) outcomes. Barschkett et al. (2022) find that increasing the early retirement age – or, put differently, spending less time in retirement – increases the likelihood of mental health diagnoses, musculoskeletal diseases and obesity. Furthermore, in a follow-up study, the authors show that healthcare costs increased due to more years in the labor force (Geyer et al., 2023).

3 Data

3.1 Data sets

We use data from two sources. The first is the German National Cohort (NAKO), a large biomedical study comparable to the UK Biobank. It is an examination that covers information on medical inspections and sociodemographic and socioeconomic factors of individuals living in Germany (Peters et al., 2022). As a second source, 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).

NAKO – The German National Cohort

The NAKO is the most extensive epidemiological examination in Germany to date, which involves the interdisciplinary collaboration of 27 German scientific institutions, including 15 universities, four Helmholtz health centers, four institutes of the Leibniz Association, and four other national research institutes. The NAKO aims to investigate the causes and the development of major chronic diseases. Starting in 2014, the NAKO examined about 205,000 individuals aged between 20 and 69 in one of their 18 different study centers throughout 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). The second wave with identical study protocol was in the field from 2018 to 2024, but the data are not yet ready to be used by scientists not involved in the data collection. Thus, we use the cross-section of the baseline examination for our analysis.

Unfortunately, the NAKO does not include information on individuals' exact birth years. Instead, we estimate the birth year by subtracting the age at examination from the respective examination calendar year. This approach introduces a potential misspecification for the 1952 birth cohort. To address this limitation, we conduct a robustness check by excluding the 1952 cohort and show that our results remain unchanged.

Administrative health insurance data

Our second data source is a panel dataset of administrative health records from 2012 to 2019. Figure A1 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, exact date of retirement entry, type of retirement, and sex.

3.2 Sample selection and retirement

We study retirement behavior and its effects on the cognition of women born three years before and after the pivotal cohort of the pension reform. Specifically, we use data on women born between 1949 and 1954 in both data sets. Table 1 shows the number of observations by age and (calculated) birth cohort for both datasets. Since the NAKO baseline study was conducted between 2014 and 2019, birth cohorts 1949–1951 are between 63 and 70 years old at the day of the assessment (60–70 in the health insurance data). Birth cohorts 1952–1954 are between 60 and 67 in the data. We use an estimation sample of women aged 63 to 67 to have the same age range covering both the cohorts affected and unaffected by the reform. This age interval leaves us with 4,788 and 845,000 observations in the unaffected group and 4,414 (667,000) observations in the affected group.

Both data sets have different but comparable information on retirement. The NAKO includes information on the labor force status, differentiating between three categories: employed, unemployed and not in the labor force. Moreover, individuals state the date of

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

	NAKO		Health insurance claims data	
Age	1949 – 1951 Not affected	1952 – 1954 Affected	1949 – 1951 Not affected	1952 – 1954 Affected
60	0	103	216,523	211,418
61	0	490	217,500	209,209
62	0	1,028	216,916	206,182
63	103	1,594	215,634	203,274
64	583	1,476	213,558	201,273
65	1,231	939	210,426	197,326
66	1,545	347	208,320	131,609
67	1,326	58	205,557	66,127
68	714	0	202,794	0
69	250	0	134,155	0
70	23	0	54,049	0
Observations				
Total	5,775	6,035	2,095,432	1,426,418
Sample 63-67	4,788	4,414	1,053,495	799,609

Notes: This table displays the number of observations by age and birth cohort in both datasets. Note that the year of birth in the health insurance data is exact, while it is derived from the age and year of interview in the NAKO. Age is age at the interview. Birth cohorts 1949-191 and 1952-1954 are summed up. The Appendix reports disaggregated data by birth cohort in Tables A2 and A3 in the Appendix.

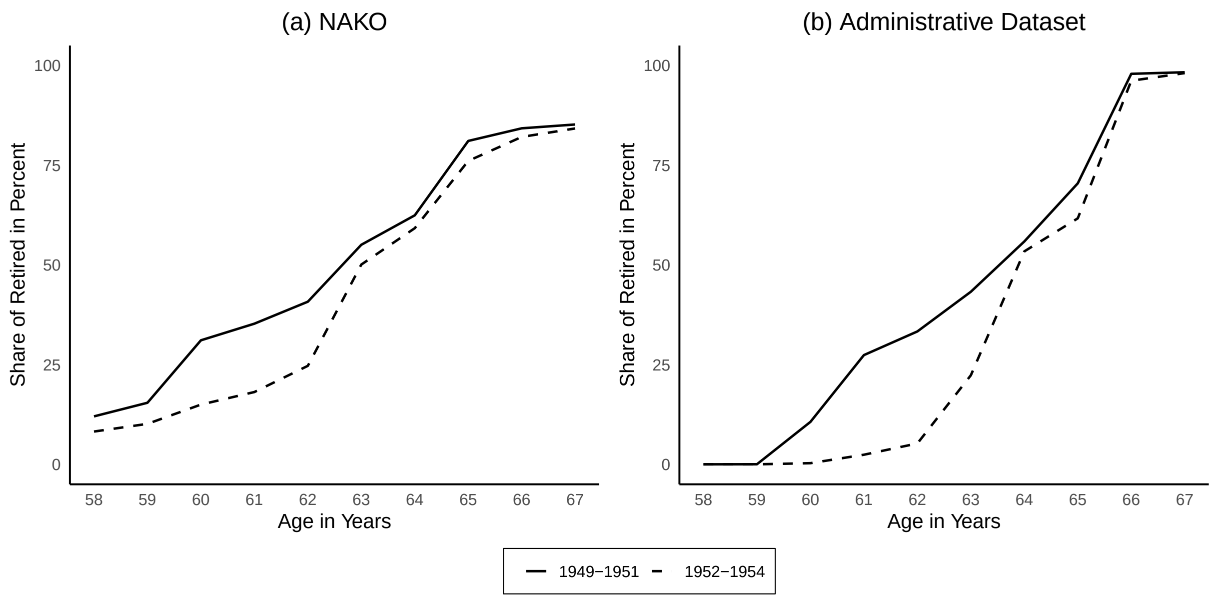
retirement if retired. We consider women retired if they declare their status as not part of the labor force and state a specific retirement age. Women who simultaneously mention a date of retirement and current part- or full-time work are not treated as retired. In the administrative dataset, we do not fully observe the labor force status, but the retirement status and the type of retirement (e.g., retirement due to age or retirement due to disability).

Since we know the retirement date, we can infer the retirement status at ages younger than 63, even though individuals in the data set are at least 63. In Figure 1, panel (a), we use the sample and the retirement information to create a pseudo-panel where women appear once per age in years with their retirement status. We observe the well-known gap in the retirement status of women between the ages 60–62 in pre- vs. post-reform cohorts. Similarly to Geyer and Welteke (2021), who show this with administrative data from public pension insurance accounts, we observe that the retirement rate is around 20 percentage points lower for women born after the pivotal cohort. In contrast, as can be expected, there are no striking differences in retirement status between cohorts younger than 60 and older than 62. Even though retirement information in the NAKO is self-reported and the birth year is derived, it is reassuring that we observe a similar reform effect on retirement as the literature. This makes us confident that the data allow to exploit the reform in order to estimate retirement effects on cognition. Note that the pseudo-panel was only generated to create Figure 1, but we use the cross-section in the regression analysis.

Panel (b) in Figure 1 also displays the share of retired women but with the administrative dataset. We see notable similarities and differences. The similarity is the large gap in retirement at ages 60 to 62 between cohorts. Differences are the shares in retirement in general. This is due to different definitions and differences in sample selection. We exclude women who have been retired for more than 10 years in both datasets. We have to assume in the NAKO that this is for reasons other than old age retirement – e.g., due to bad health. Nevertheless, this is an imperfect approach, and we still have around 20 percent of women classified as retired even though they do not qualify for old-age pensions.[4] Moreover, we allow women to be non-retired even at ages older than 67 if they report to have a job. In the health insurance data, we exclude individuals who explicitly mention other reasons for retirement than age, for example, due to a health-related inability to work. Moreover, we do not observe if women have an additional job besides retirement, leading to a retirement rate of almost 100 percent at ages older than 66.

Our explanatory variable of interest in the analysis will be the retirement duration, measured in years. It is calculated as a current date minus the date of retirement. We use the gap in retirement due to the reform as an identifying variation to estimate the effects of retirement duration on cognitive abilities.

Figure 1: Differences in labor force status between birth cohorts



Note: Panel (a) shows the labor force status of women by their age and year of birth based on an artificial panel constructed with the NAKO. Panel (b) shows the share of retired individuals in our administrative dataset. The solid and dotted lines correspond to women affected and unaffected by the reform, respectively. Information on 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]This is not a problem for the estimation results. These 20 percent may be termed always-takers who anyway do not contribute to the estimated local average treatment effects we estimate below.

3.3 Outcome variables

Cognitive abilities are 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” (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 to capture memory and processing speed, crystallized intelligence measures abilities and knowledge 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 three cognitive scores representing our outcome measures.[5] We use word recall, semantic fluency, and the Stroop effect. These neuropsychological tests aim to assess verbal episodic memory, processing speed, or executive control.[6] From the insurance claims data we include diagnoses of MCI and dementia. The dataset is on individual-year level, with our outcome variable equal to 1 if an individual is diagnosed in the current year.

Word recall

In step 2 of the NAKO test battery, after other health assessments, 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 was presented for two seconds. 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. 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, gives our first measure of cognitive abilities in the application. 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 a measure of fluid intelligence.

Semantic fluency

Participants were asked to name as many animals as possible within one minute. The number of animals mentioned provides the measure of semantic fluency. Schiele and

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

[6]See Figure 2 and Table 2 for summary statistics of outcome variables and individual characteristics.

Schmitz (2023) use the number of animals mentioned to measure verbal fluency in examining the effect of health shocks on cognitive abilities. Freise et al. (2022) use verbal fluency to investigate the effects of late-career unemployment on cognition. Verbal or semantic fluency is mainly present in neuropsychological research, specifically on Alzheimer's disease or dementia, e.g., Murphy et al. (2006), Mueller et al. (2015), and Nikolai et al. (2018). Semantic fluency is considered 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 Stroop effect effect quantifies the temporal difference between two Stroop task tests carried out in the test battery. 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 is calculated as the time difference between the second and first tests. 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 potential effects consistent with the other cognitive measures. The Stroop effect measures 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 and a more complex Stroop task shows whether a person has better control over response inhibition. Different works on the association of aging, mental diseases, and cognitive functioning rely on the Stroop test, for instance, Spieler et al. (1996) or West and Alain (2000). These works are predominantly located in the field of neuroscience.

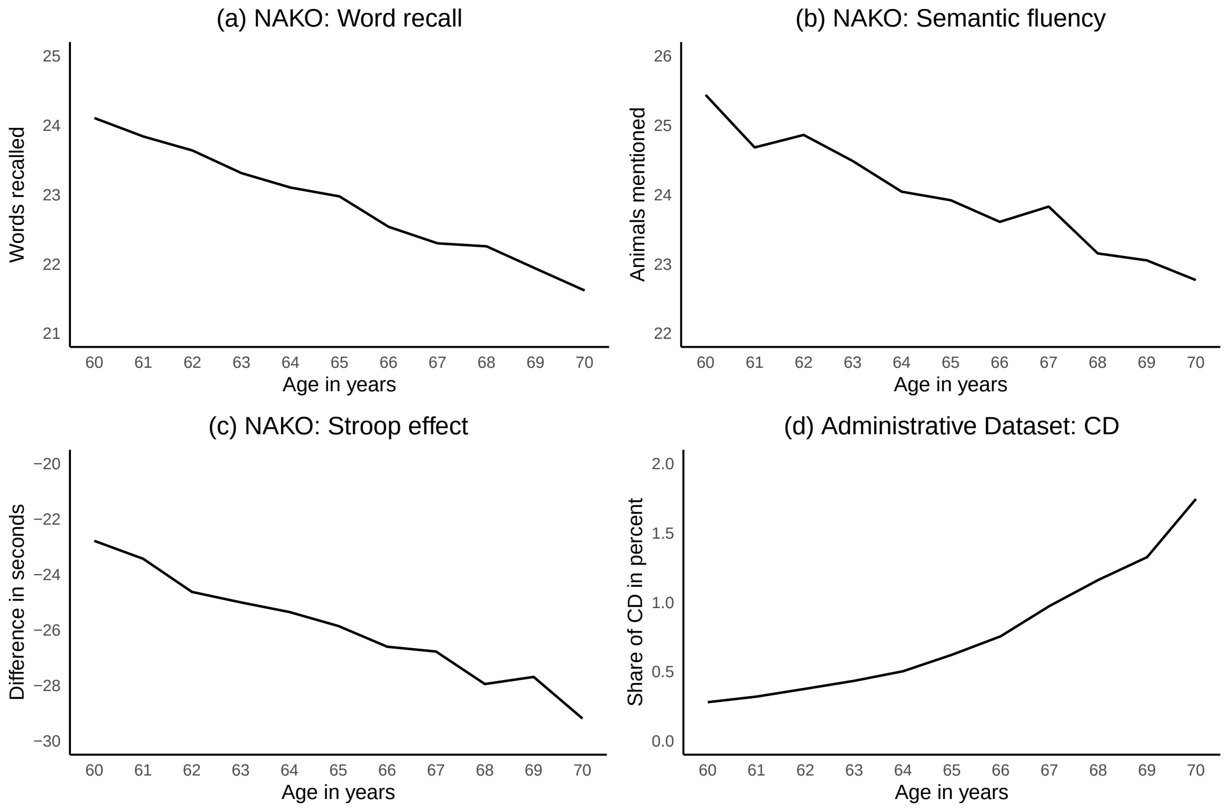
Administrative data: Diagnoses of cognitive disorders

We use all ICD-10-GM codes listed in Table A1 in the Appendix to identify dementia and mild cognitive impairment in our data as diagnoses of cognitive disease (CD). Since individuals born around the reform cutoff are 67 years old in 2019 we do not only rely on the use of dementia for our analysis as incidence is still very low at that age. Instead, we add the preliminary diagnosis of MCI and combined it with dementia, resulting in our pooled medical outcome for diagnoses of CD. Put differently, our outcome variable is 1 if a women is diagnosed with any ICD-10-GM code for MCI or dementia. 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 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 using 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 Descriptive statistics

We plot our outcome measures of cognitive abilities and CD diagnoses by age and reform exposure in Figure 2. All outcomes of cognitive abilities show a clear trend with increasing age. Cognitive abilities decline while the share of CD diagnosis increases.

Figure 2: Outcome variables by age



Note: This figures show means of the outcome variables in our sample by age and year of birth. Diagnoses are defined as shown in Table A1.

Table 2 presents descriptive statistics for both datasets separately by reform eligibility. As expected, average retirement duration is shorter in the treated group (1.3 and 2 years) compared to the untreated group (2.5 and 4 years). Figure 1 verifies this observation and reveals that individuals in the treated cohorts tend to retire at a later age than those in the untreated. However, it is noteworthy that the average age at retirement is lower in the treated group (62.1 years) compared to the untreated group (62.2) in the NAKO. The reason is that this number is conditional on being retired. Since more women in the treated group are not yet retired, more weight is given – in calculating this descriptive statistic – to women who retired for other reasons than age and at younger ages in the treated group. Both Figure 1 and the first stage regression results show that – conditional on current age – women in the treated group have higher retirement ages also in the NAKO data. This discrepancy is not observed in the administrative data because women who retired due to other reasons than age are excluded.

Table 2: Descriptive statistics by reform eligibility

	Born 1949-1951 Not affected		**Born 1952-1954** Affected	
	Mean	(SD)	Mean	(SD)
NAKO				
Socioeconomic factors				
Age	65.72	(1.06)	64.05	(1.00)
Birth year	1950.2	(0.79)	1952.8	(0.77)
Retired (in %)	86.81		65.58	
Age at retirement (if retired)	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)
Cognition measures				
Word recall	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:	4,788		4,414	
Insurance claims data				
Socioeconomic factors				
Age	64.98	(1.41)	64.57	(1.26)
Birth year	1950.0	(0.81)	1952.8	(0.80)
Retired (in %)	72.8		58.28	
Age at retirement (if retired)	63.19	(2.17)	64.14	(1.42)
Years retired	2.54	(2.27)	1.32	(1.45)
Cognitive diagnoses				
Cognitive Disorder (in %)	0.66		0.59	
Dementia (in %)	0.5		0.41	
MCI (in %)	0.19		0.21	
Observations:	1,053,495		799,609	

Note: This table shows descriptive statistics for women in the selected samples.

The prevalence of dementia in the administrative dataset is 0.5% in the untreated group while the prevalence of MCI is lower at 0.19%, given its role as a preliminary diagnosis for dementia. The low prevalence in the dataset could also reflect underreporting in the administrative data, possibly due to mild symptoms in the relatively young age group preventing individuals from seeking medical help or a tendency among physicians not to diagnose MCI. The prevalence of diagnoses in our administrative dataset, along with MCI often serving as a preliminary diagnosis for dementia, underscores the need to use the pooled outcome measure CD. With such a small number of individuals diagnosed with each condition, estimates are prone to high noise due to limited variation in the outcome variable. Pooling the outcomes reduces this noise, resulting in more stable and precise estimates.

4 Estimation strategy

As a benchmark, we run the following OLS regressions:

$$Y_i = \quad \rho^{OLS}\ YearsRetired_i + \gamma_a + \lambda_c + X_i'\beta + \varepsilon_i \tag{1}$$

The dependent variable, Y_i, denotes measures of cognitive abilities (word recall, semantic fluency, and the Stroop effect) and CD of women i. $YearsRetired_i$ is the retirement duration measured in years. The parameter ρ^{OLS} captures the association between retirement duration and cognitive abilities. Furthermore, we control for regional (λ_c) and age in years (γ_a) fixed effects. The vector X_i may include further variables depending on the specification and employed data, such as an indicator for the birthplace of Germany.[7]

One potential threat to a causal interpretation of the OLS approach is self-selection into retirement: individuals with cognitive problems may choose or even have to retire such that they are observed to be retired longer. To address this problem and estimate a causal effect of retirement duration, we implement a two-stage least-squares model (2SLS) exploiting the pension 1999 reform. This builds on Geyer and Welteke (2021) and Geyer et al. (2020), who investigate the labor market effects of the reform, documenting the relevance of our instrument.

Specifically, we instrument years retired with a dummy indicating women born in 1952 or after. For them, the right to claim public pension benefits before age 63 was revoked. $Reform_i$ acts as an instrumental variable and is expressed by the indicator function

[7] We control for birthplace since our measures of cognitive abilities are highly correlated to the proficiency of speaking German (see Kleinedamm et al., 2022). While individuals 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, it is not essential to control for this variable, as it is uncorrelated with our instrument. However, we may potentially gain some precision. Hence, it is not a problem that we do not observe birthplace in our administrative dataset.

$Reform_i = \mathbb{1}(BirthCohort_i \geq 1952)$. Equations (2) and (3) refer to the first and second stages, respectively, with coefficients indexed by FS and $SeSt$:

$$YearsRetired_i = \alpha\ Reform_i \qquad + \gamma_a^{FS} + \lambda_c^{FS} \quad + X_i'\beta^{FS} + \varepsilon_i^{FS} \tag{2}$$

$$Y_i = \rho\ \widehat{YearsRetired}_i + \gamma_a^{SeSt} + \lambda_c^{SeSt} + X_i'\beta^{SeSt} + \varepsilon_i^{SeSt} \tag{3}$$

In the first-stage equation, α is the reform effect on retirement duration. In contrast, ρ is the effect of the reform on the cognitive outcomes, (*i*) for individuals who have to retire later due to the reform, and (*ii*) normalized to a one-year change in the retirement duration. In short, this is the local average treatment effect (LATE). To interpret α and ρ as causal, and more specifically, ρ additionally as a LATE, specific assumptions on the instrument need to be fulfilled.

First, the reform needs to be exogenous. The date of an individual's birth is not within their control; the reform applies to every woman, and it is unlikely that women born shortly before and after 1 January 1952 differ for other reasons. Hence, this assumption seems plausible. Second, it is necessary to assume that the reform's sole impact on individuals' cognitive abilities is via its effect on retirement duration (i.e., the exclusion restriction). Third, the pension reform should significantly affect the number of retirement years – a property frequently demonstrated in the literature and verified in our data in Figure 1. Fourth, it is necessary to assume the absence of defiers. That is to say, women who retire earlier if affected by the reform than they would have preferred without the reform do not exist. This is unlikely a strong assumption due to the stark financial incentives. Assuming these conditions are met, ρ gives the average effect of retirement for women affected by the reform (the compliers) adjusted to a one-year increase in retirement duration. Note that if one were to interpret this effect as the complier's marginal effect of retirement duration on cognitive abilities, one would need to additionally assume homogeneous treatment effects along retirement duration (which we are not willing to take).

Potential threats to the identification strategy

Examining women born a few years apart raises concerns about potential cohort effects driving our results. Birth cohorts differ in initial conditions, 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 at a wide range of birth years from 1934 to 1956. In their first-differencing approach, the authors define birth year groups of five years, thereby demonstrating robustness to possible effect heterogeneities. Combined with our empirical approach, our samples do not allow for a proper separation of age and cohort effects. Since NAKO is a cross-sectional dataset observed at different time points, age is always observed at different birth years. It is, therefore, defined as a combination of birth year and interview year. As our sample consists of only six different birth years, we are confident that potential cohort effects represent a minor threat to our estimation strategy. Previous literature usually looks at a broader set of birth years when investigating cohort effects (Gerstorf et al., 2011; Zelinski and Kennison, 2007; Karlsson et al., 2015). However, we provide a robustness check against cohort effects by controlling for linear cohort trends on both sides of the cutoff at birth year 1952.

As mentioned above, the NAKO does not provide information regarding the exact year of birth. We calculate the birth year from the difference between age and interview year. Potentially, this may affect the reform assignment. If a person's birthday in a specific calendar year is after the interview date, there is a possibility of incorrectly assigning exposure to the reform for the 1952 birth cohort. We will address this by excluding individuals born in 1952 from our baseline regression as a robustness check.[8] We further check for robustness to wave effects, confounding by a concurrent compulsory schooling reform, and whether the results change when our sample is extended by age. Also, we conduct placebo tests with a hypothetical reform cutoff.

5 Results

5.1 Main results

OLS results, first-stage estimates, and 2SLS results for the different outcomes are reported in Table 3. The OLS results show significant correlations between retirement duration and cognitive functioning. An additional year in retirement is associated with a decrease in words recalled by 0.108. Estimates of semantic fluency show a negative correlation of -0.118, while the Stroop effect exhibits a negative correlation of -0.205. An additional year spent in retirement is associated with a 0.02 percentage point higher risk of being diagnosed with CD. Nevertheless, since retirement duration is endogenous, OLS results do not offer insights into the effects on cognitive functioning.

The 2SLS results provide evidence of a negative local average treatment effect of retirement duration on cognitive abilities for women who change their retirement entry due to the

[8] If the assumption that all individuals had their birthday before the interview date is violated, their true birth year would shift to the next older cohort. The 1952 cohort is particularly exposed to this, as it is the only cohort where the eligibility status for the reform would change. Therefore, there is no need to drop other cohorts.

Table 3: Main results - OLS and 2SLS

	NAKO data			Insurance claims data
	Word recall (1)	Semantic fluency (2)	Stroop effect (3)	Diagnosis of CD (in %) (4)
OLS:				
Years retired	-0.108^{***}	-0.118^{***}	-0.205^{***}	0.02^{***}
	(0.021)	(0.026)	(0.057)	(0.004)
Reduced form:				
Reform	0.368^{***}	0.220	0.981^{***}	−0.01
	(0.131)	(0.165)	(0.363)	(0.012)
First stage:				
Reform	-0.591^{***}	-0.582^{***}	-0.586^{***}	-0.881^{***}
	(0.064)	(0.064)	(0.064)	(0.002)
2SLS:				
Years retired	-0.623^{***}	−0.379	-1.672^{***}	0.011
	(0.228)	(0.283)	(0.642)	(0.013)
Placebo reform: 1948-1949 vs. 1950-1951. Reduced form:				
Reform	−0.074	0.002	−0.143	−0.007
	(0.142)	(0.177)	(0.395)	(0.017)
Mean of DV:	22.93	24.05	−25.66	0.628
	(4.78)	(6.33)	(13.39)	(7.899)
Observations	9,283	9,382	9,280	1,853,104

Note: This table shows coefficients for the outcome variables of cognitive abilities and CD. The first row shows the estimated coefficients from an OLS regression based on Equation (1). The second and third row show the results from the reduced form and the first stage based on Equations (2). The fourth row shows results from the corresponding 2SLS estimation from Equation (3). The fifth row shows results from a reduced form analysis using 1950 as a placebo cutoff comparing cohorts 1948-1949 to cohorts 1950-1951. The regressions in columns 1 to 3 include a dummy for the birthplace of Germany. All regressions include age and regional fixed effects. Regional fixed effects are on the study center level for columns 1 to 3 and the municipality level for column 4. OLS and 2SLS coefficients in column 4 are multiplied by 100 for better readability. Numbers of observations refer to regressions in rows 1–4. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

pension reform. The first stage effect is highly significant in both datasets giving credibility to the strength of our instrument. Looking at the 2SLS results reveals that the direction of the second-stage effects corresponds with the OLS results. We find effects that indicate a decline in cognitive abilities due to an increasing duration of retirement. An additional year in retirement significantly lowers the words recalled by 0.62 words, which is 3% of the sample mean and 13% of the SD. Results for the Stroop effect confirm these findings. An additional year in retirement increases the time difference between the two Stroop tests by 1.67 seconds, which is 6% of the sample mean and 12% of a SD. Even though estimates for semantic fluency are not statistically significant they correspond to the other estimates and indicate a negative relationship. Regarding the effect of years in retirement on the risk of being diagnosed with CD we only find a very small and statistically insignificant effect of an increase in the risk of being diagnosed by 0.011 percentage points by the year of retirement. This equals 2 % of the sample mean risk of a CD diagnosis. Thus, while extremely small in absolute terms, the relative size of the effect is not too far away from the estimated effects in columns 1 to 3.

Our results indicate that an additional year in retirement causes a significant decline in women's cognitive functioning. 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. Yet, their effect size is smaller than 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 align with literature in neuropsychological research that portrays age-specific patterns in different intelligence fields during an individual's lifespan (Rohwedder and Willis, 2010; McArdle et al., 2002). Semantic fluency, expressed by enumerating animal names, corresponds to crystallized intelligence, whereas word recall and the Stroop effect show fluid intelligence. Our findings appear reasonable since fluid intelligence is exposed to a more substantial decline than crystallized intelligence.

In the lowest set of results in Table 3 we report results of a placebo regression. Here, we set the year 1950 as a hypothetical reform cutoff and compare cohorts 1948–1949 with 1950–1951 to ensure that all cohorts are not affected by the actual reform. Note that we cannot go further back in birth cohorts as these are hardly filled in the NAKO that stops data collection at age 69, leaving only a handful of observations in birth cohorts 1947 and before. Again, we only include ages that are present in both the treated and untreated groups. These are now the ages 66 to 69. Significant effects on the outcome measures disappear completely and the coefficients of the reduced form clearly move towards zero. This makes us confident that significant positive effects in the reduced form of our main specification are driven by changes in retirement regulations.

5.2 Robustness checks

5.2.1 Cohort effects

Cognitive ability in later life may be affected by cohort effects. This could be due to differences in early life environments such as health, education, or socio-economic status (Case and Paxson, 2010; Currie, 2009). However, as our analysis is based on cross-sectional data observed at different points in time, we cannot fully disentangle age and cohort effects.

Despite this limitation, we are confident that our findings—derived from a sample of only six birth cohorts—are not substantially driven by potential cohort effects. Nonetheless, it is essential to acknowledge this issue and provide further evidence of robustness. As part of our first robustness check, we control for a linear cohort trend that allows for different slopes at each side of the cutoff, as defined by the birth year 1952. We extend Equation (3) with a linear birth cohort trend $\widetilde{BirthYear}_i$ centered at the pivotal cohort of 1952. Since a cohort trend is highly correlated with age and the IV, we reduce overall collinearity in this specification by running the regression without age fixed effects. Equation (4) displays this approach with an OLS regression:

$$Y_i = \rho\ YearsRetired_i + \delta\ \widetilde{BirthYear}_i + \theta \cdot Reform_i \times \widetilde{BirthYear}_i + \lambda_c + X_i'\beta + \varepsilon_i \tag{4}$$

The corresponding first stage is adjusted accordingly. Table 4 shows the results of our robustness check.

Table 4: Cohort effects – 2SLS

	NAKO data			Insurance claims data
	Word recall (1)	Semantic fluency (2)	Stroop effect (3)	Diagnosis of CD (in %) (4)
Reduced form: Cohort effects				
Reform	0.452**	−0.006	0.831	−0.0002
	(0.197)	(0.251)	(0.563)	(0.0002)
First stage: Cohort effects				
Reform	−0.406***	−0.401***	−0.405***	−0.762***
	(0.107)	(0.106)	(0.106)	(0.006)
2SLS: Cohort effects				
Years retired	−1.113**	0.014	−2.050	0.021
	(0.548)	(0.626)	(1.468)	(0.032)
Mean of DV:	22.93	24.05	−25.66	0.628
	(4.78)	(6.33)	(13.39)	(7.899)
Observations	9,283	9,382	9,280	1,853,104

Note: This table shows coefficients for a 2SLS estimation based on Equation (4). First, results from the first stage are presented, followed by results from the corresponding 2SLS estimation. The regressions in columns 1 to 3 include a dummy for the birthplace of Germany. All presented regressions include regional fixed effects on the study center level for columns 1 to 3 and on the municipality level for column 4. OLS and 2SLS coefficients in column 4 are multiplied by 100 for better readability. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Except for semantic fluency, the reduced form effects are hardly changed compared to the baseline specification. Yet, the standard errors increase. The effect of retirement on semantic fluency is zero in this specification. Important differences compared to above are the increased 2SLS effects for word recall and the Stroop test. Apparently, this is due to a smaller first stage. Changes in the estimates for CD diagnosis are unremarkable.

Although the effects of fluid intelligence measures (word recall, Stroop test) are larger, we argue that this specification is not our preferred one. The (first stage) reform effect on retirement duration is probably underestimated. The regression specification may exacerbate the influence of the misspecification of the 1952 birth cohort because the cohort trend assigns more weight to the cutoff, i.e., 1951 vs. 52 comparisons. A quick back-on-the-envelope calculation also shows this underestimation. While there is no estimate in the literature on the effect of the pension reform on retirement duration, Geyer and Welteke (2021) find that 60-62-year-old women affected by the reform have a 27.6 percentage points lower likelihood of receiving old-age pensions. For women aged 63 and older, this translates into an average reduced retirement duration of three times 0.276 = 0.828 years, comparable to the 0.881 first stage using the insurance claims data in Table 3. Since the basis of the retirement definition in the NAKO is the status out of the labor force, the first stage results compare best to Geyer and Welteke (2021)'s combined effects on old-age pension receipt, disability pension and their "residual category". This would be a three times 0.221 = 0.663 years effect, much closer to our first stage in Table 3 than in Table 4. Thus, we believe that the first stage in the baseline specification is close to results that use detailed data from the German pension insurance while it is underestimated when we control for cohort effects. We therefore proceed with the estimates from Equation (3) as our preferred specification, as we believe that they more closely correspond to the true effect.

5.2.2 Further robustness checks

Misspecification of birth year

As discussed, the NAKO 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 affected by the reform, even though they may not be. As a result, this could lead to underestimating the first-stage effect in absolute terms and underestimating 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.

We address this issue by excluding women born in 1952 from our sample. This approach allows us to distinguish between individuals exposed and unexposed to the reform more accurately. However, a significant drawback is the loss of statistical power, as excluding this cohort reduces the sample size by approximately 20 percent and necessarily also excludes correctly classified individuals. 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 5. The first row repeats the 2SLS estimates obtained in Table 3. The second row shows estimates obtained with the revised sample, excluding the 1952 cohort. As can be seen, the effects for the Stroop effect are larger and remain statistically significant. Effects on word recall are smaller and lose significance. However, effects are statistically significant at the 10%-level and indicate a negative effect. Effects on CD increase and turn significant at the 10% level, which is surprising given that the birth year is correctly specified in the insurance claims data. In general, negative effects persist and do not suggest a potential bias due to a misspecification of birth year.

Table 5: Robustness checks

	NAKO data			Insurance claims data
	Word recall (1)	Semantic fluency (2)	Stroop effect (3)	Diagnosis of CD (in %) (4)
2SLS regression results:				
Main results				
Years retired	-0.623^{***}	-0.379	-1.672^{***}	0.011
	(0.228)	(0.283)	(0.642)	(0.013)
Without 1952 cohort				
Years retired	-0.467^{*}	-0.488	-2.033^{***}	0.026^{*}
	(0.255)	(0.322)	(0.743)	(0.015)
Without waves 2014 and 2019				
Years retired	-0.616^{**}	-0.329	-1.746^{**}	-0.016
	(0.257)	(0.324)	(0.735)	(0.016)
Sample expansion: Age 60 to 67				
Years retired	-0.582^{***}	-0.344	-1.620^{***}	0.011
	(0.220)	(0.274)	(0.622)	(0.012)
Without 1953 & 1954 cohorts				
Years retired	-0.928^{***}	-0.381	-1.859^{**}	-0.011
	(0.329)	(0.395)	(0.880)	(0.02)

Note: This table shows coefficients for the outcome variables of cognitive abilities. 2SLS results of Table 3 are shown in the first row. In the second row, individuals born in 1952 are excluded as a robustness check for the calculated birth year. In the third row, waves 2014 and 2019 are excluded to show robustness against wave effects. In the fourth row ages 60 to 67 are included. In the fifth row, cohorts 1953 and 1954 are excluded. The observations without the 1952 cohort range between 7,289 and 7,293 in NAKO and take 1,516,073 in the administrative dataset. The numbers of observations without waves 2014 and 2019 range between 8,766 and 8,770 in NAKO, while the number in the administrative data takes 1,443,280. The numbers of observations including ages 60 to 67 range between 10,913 and 11.034 in NAKO and take 3,130,852 in the administrative dataset. The numbers of observations without cohorts 1953 and 1954 range between 6,823 and 6,892 in NAKO and 1,390,523 in the administrative dataset. The regressions in columns 1 to 3 include a dummy for the birthplace of Germany. All presented regressions include age and regional fixed effects. Regional fixed effects are on the study center level for columns 1 to 3 and the municipality level for column 5. OLS and 2SLS coefficients in column 4 are multiplied by 100 for better readability. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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 in row three of Table 5 show that second-stage estimates remain consistent with those in Table 3, with no significant changes.

Sample expansion by age

The main reason for restricting our sample to ages 63-67 is that these are the ages that we observe for both treated and untreated women in NAKO. However, the administrative data allows for analyzing a wider age range, as shown in Table 1. We therefore expand the sample by including ages 60 to 67 and show 2SLS results in row five of Table 5. Estimates show that effects remain unchanged for measures of cognitive abilities and the risk of being diagnosed with CD. This is unsurprising since we expect effects to be particularly pronounced in older ages.

Compulsory schooling reform

In our sample, we include individuals born up to 1954, while Geyer et al. (2023) limit their sample to those born up to 1952, as individuals born after this year were affected by a compulsory schooling reform that had positive health effects (Kemptner et al., 2011). In our final robustness check, we align with Geyer et al. (2023) by restricting the sample to individuals born between 1949 and 1952. Table 5 shows that the estimated effects for all outcomes remain robust in both size and significance, consistent with our main results. This finding is supported by literature (Kamhöfer and Schmitz, 2016; Seblova et al., 2020). For instance, Kamhöfer and Schmitz (2016) found no impact of the compulsory schooling reform on cognitive outcomes, and Seblova et al. (2020) reported no effect of a similar reform in Sweden on dementia risk.

6 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.

Our findings indicate that an additional year in retirement has a significant negative effect on the cognitive abilities of women whose preferred retirement is deferred to the age of 63 due to the pension reform. This holds in particular with respect to cognition measures related to fluid intelligence. Effects are smaller and not statistically significant for a measure of crystallized intelligence. Moreover, up to the age of 67, we find no significant effect of retirement duration on the risk of being diagnosed with a cognitive disease, which includes conditions such as dementia and MCI. However, such effects may emerge at older ages, suggesting that the relationship between retirement and cognitive decline may become more pronounced later in life.

Our findings contribute to the literature by focusing the analysis on a precise group in a single country: women in Germany. We exploit a substantial retirement reform that induced a pronounced change in women's early retirement age and do not use, for instance, country-specific fixed retirement ages. While our setting has the benefit of a high internal validity, it may come at the cost of a limited external validity. For instance, we cannot extrapolate the effects to other countries, males, or different hypothetical retirement ages or durations (as, for instance, Schmitz and Westphal, 2021, who, however, need to impose stronger assumptions for internal validity). Notably, our results are different from the general health effects of the reform under study (Barschkett et al., 2022; Geyer et al., 2023), making clear the importance of including multidimensional outcomes when retirement effects are analyzed.

Of course, our study has limitations. One key limitation is the inability to observe long-term effects, as women born around 1952 are just entering their 70s. Yet, the identified effects on measures of fluid intelligence might be a precursor of diagnosed cognitive impairment some years later. Ultimately, based on our results, we conclude that extending the retirement entry to a later date could potentially mitigate the decline of cognitive abilities induced by additional years in retirement.

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.

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.

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.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, 47(1):87–122.

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.

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).

Kamhöfer, D. A. and Schmitz, H. (2016). Reanalyzing zero returns to education in germany. *Journal of Applied Econometrics*, 31(5):912–919.

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.

Kemptner, D., Jürges, H., and Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: evidence from Germany. *Journal of Health Economics*, 30(2):340–354.

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.

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.

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.

Seblova, D., Fischer, M., Fors, S., Johnell, K., Karlsson, M., Nilsson, T., Svensson, A. C., Lövdén, M., and Lager, A. (2020). Does Prolonged Education Causally Affect Dementia Risk When Adult Socioeconomic Status Is Not Altered? A Swedish Natural Experiment in 1.3 Million Individuals. *American Journal of Epidemiology*, 190(5):817–826.

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.

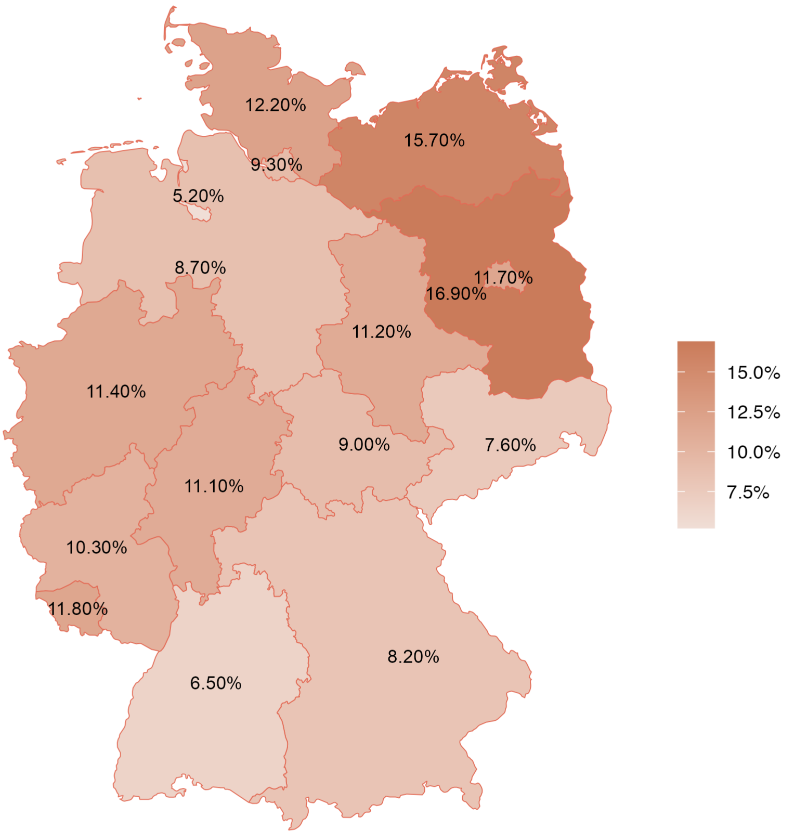
West, R. and Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, 37(2):179–189.

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: Regional distribution of insured individuals



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

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: 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	104	0	104
61	0	0	0	0	115	378	0	493
62	0	0	0	93	418	534	0	1,045
63	0	0	105	471	605	532	105	1,608
64	0	120	467	626	535	332	587	1,493
65	113	466	656	546	339	63	1,235	948
66	460	571	539	293	58	0	1,570	351
67	556	466	316	58	0	0	1,338	58
68	426	244	59	0	0	0	729	0
69	206	46	0	0	0	0	252	0
70	23	0	0	0	0	0	23	0
Observations								
Total	1,784	1,913	2,142	2,087	2,070	1,943	5,839	6,100
Sample 63-67	1,129	1,623	2,083	1,994	1,537	927	4,835	4,458

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

Table A3: Observations by age and birth year in administrative dataset

	birth year						Reform	
Age	1949	1950	1951	1952	1953	1954	Not affected	Affected
60	71,318	72,977	72,228	71,401	70,138	69,879	216,523	211,418
61	72,136	72,969	72,395	70,841	69,479	68,889	217,500	209,209
62	71,956	72,986	71,974	70,103	68,429	67,650	216,916	206,182
63	71,917	72,465	71,252	69,160	67,389	66,725	215,634	203,274
64	71,407	71,840	70,311	68,295	66,763	66,215	213,558	201,273
65	70,567	70,721	69,138	66,866	65,448	65,012	210,426	197,326
66	69,942	69,884	68,494	66,583	65,026	0	208,320	131,609
67	68,845	69,036	67,676	66,127	0	0	205,557	66,127
68	67,883	68,108	66,803	0	0	0	202,794	0
69	66,978	67,177	0	0	0	0	134,155	0
70	54,049	0	0	0	0	0	54,049	0
Observations								
Total	756,998	708,163	630,271	549,376	472,672	404,370	2,095,432	1,426,418
Sample 63-67	352,678	353,946	346,871	337,031	264,626	197,252	1,053,495	799,609

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