Radiation and Human Capital: Long-run Evidence from Exposure outside the Womb

Benjamin Elsner
IZA, UCD Geary Institute, and CReAM

Florian Wozny
IZA

15 November 2017

Abstract: This paper studies the impact of long-run exposure to radiation on cognitive skills. We focus on Germany, which received a large amount of fallout after the Chernobyl nuclear disaster in 1986, resulting in a permanent increase in radiation levels in the soil. For identification, we exploit the unequal distribution of the fallout due to idiosyncratic rainfall, and the fact that the population had no knowledge of the local level of contamination. Our results show that people who lived in a more contaminated area in 1986 perform significantly worse in standardized cognitive tests 25 years later. This effect is driven by older cohorts (aged 10-35 in 1986), while we find no effect for people who were first exposed during early childhood. These findings suggest that radiation accelerates the decline in cognitive skills at older ages.

JEL-Classification: J24, Q53

Keywords: *Environment, Human Capital, Radioactivity, Cognitive Skills*

We would like to thank Hoyt Bleakley, Leonardo Bursztyn, Deborah Cobb-Clark, Joan Costa-i-Font, Thomas DeLeire, Olivier Deschênes, Jon Guryan, Adam Isen, Shelly Lundberg, as well as audiences at IZA, U Freiburg, the 5th IZA Workshop on Environment and Labor Markets, RWI Essen and the 2nd IZA Workshop on the Economics of Education for helpful comments.

Corresponding author: Benjamin Elsner, IZA, CReAM and the UCD Geary Institute. Institute of Labor Economics, Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany. Phone, +49 (228) 3894-522, Fax +49 (228) 3894-510, E-mail elsner@iza.org

Florian Wozny, Institute of Labor Economics (IZA), E-mail wozny@iza.org.

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Adults, $doi : 10.5157/NEPS : SC6 : 8.0.0$. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

1 Introduction

Pollution is known to have long-lasting consequences for people's life outcomes. Numerous studies document negative effects of air pollution, contaminated water, radiation, or other sources on birth weight, health, educational attainment, and wages. The primary identification strategy for uncovering these long-term effects is to exploit an environmental shock and compare the outcomes of people exposed while in the womb to those of a suitable control group (Almond and Currie, 2011). However, while this approach has delivered important insights about the negative effects of pollution, it is unclear whether the results generalize to the broader population, namely all people who were *not* in the womb at the time of the shock. In this paper, we provide complementary evidence by showing that pollution can even have negative effects when people are first exposed during adolescence or adulthood.

We focus on the effect of radiation, a pollutant to which every person is exposed at all times, although the degree of exposure varies between places and lifestyles. We exploit the nuclear disaster in Chernobyl/Ukraine in 1986, which led to contamination of the soil in large parts of Europe. With a half-life of over 30 years, the fallout led to a quasi-permanent increase in radiation levels in most —but not all —regions. Using geo-coded survey data from Germany, we study the impact of this contamination on people's cognitive skills more than 25 after the accident.

Our study is based the National Educational Panel Study (NEPS), which covers people born between 1952 and 1986. The NEPS offers two features that are key to our analysis. First, for every respondent, a detailed residential history is available, allowing us to link personal information with radiation data in the respondent's place of residence at the time of the disaster. Second, the survey includes eight standardized cognitive tests that were taken by the respondents when they were between 24 and 58 years old. This makes NEPS one of the few datasets with information on cognitive skills after school-leaving age.

To identify a causal effect, we exploit that the level of fallout in a region critically depended on rainfall in a 10-day window after the disaster. Places with heavy rainfall in this period received significantly more fallout than those with little rainfall. Moreover, at the time, the German population received fairly sparse information about the disaster and its consequences. Due to the sluggish release of information from the Soviet Union, people only learned about the disaster after the radioactive rain had fallen. A detailed map of radiation levels in small geographic areas was only released five years later, making it unlikely that people could avoid exposure by moving to a less contaminated area. Nonetheless, balancing tests suggest that residential sorting before 1986 is correlated with post-Chernobyl radiation levels, mainly the result of people with higher education living in cities whose geography makes them less prone to radiation. However, once we condition on the geographic determinants of receiving fallout such as altitude and average rainfall, there is no systematic correlation between observable characteristics and radiation levels.

We find significant negative effects of radiation exposure on cognitive skills 25 years after the disaster. A one-standard-deviation higher exposure to radiation reduces cognitive test scores 25 years later by five percent of a standard deviation. Because our reduced-form regression identifies a combination of biological effects and compensating behaviors, this estimate is likely a lower bound to the pure biological effect of radiation on cognitive skills. We also find the effect to be stronger for skills relying on crystallized intelligence, such as maths, reading comprehension and logical reasoning, and weaker for those relying on fluid intelligence such as scientific knowledge or IT skills. These results are consistent with the medical literature, which suggests that radiation exposure mainly affects the hippocampus, the part of the brain that governs crystallized intelligence (Monje and Dietrich, 2012).

We further study whether the effect of radiation differs between demographic groups. While we find no differential effect between men and women, we find significant differences by age of first exposure. For people exposed during early childhood (between zero and 10 years of age in 1986), we find no effect, while we find strong negative effects for older cohorts. At first glance, this result appears runs counter to much of the literature showing that pollution matters when people are exposed in the womb or during early childhood. One explanation for this finding is that the effects of radiation only materialize in older age, and the younger cohorts in our sample are simply too young to experience cognitive decline due to radiation. An additional interesting result is a stronger effect for people living in East Germany in 1986 compared to those living in West Germany. This finding is consistent with East Germans receiving little information about the disaster and, thus, having little chance to engage in compensating behavior.

In extensive robustness checks, we carefully address challenges to identification and inference. To address potential unobserved heterogeneity as well as measurement error, we use deviations from the average rainfall in the 10-day window in May 1986 as an instrument for the ground deposition. In this case, the estimates are larger, suggesting that measurement error attenuates our OLS estimates. Furthermore, a bounding analysis proposed by Oster (2016) shows that the results are unlikely explained by selection on unobservable characteristics. To assess our inference, we perform a permutation test, which reveals that our estimate is unlikely to appear by chance. Moreover, our inference is robust to multiple hypothesis corrections, which increases the confidence in our results.

This paper fills an important gap in the literature on pollution and human capital. Most of the existing evidence centers around two types of effects. One strand of literature explores the long-run effects of exposure to pollution early in life on outcomes many years later. Another strand focuses on adolescents and adults, and estimates how their exposure to pollution affects their test scores or performance at work in the short term —often on the same day. There is little evidence, however, on the long-run effects of pollution when people are exposed during adolescence or adulthood. Our research setting, combined with the availability of cognitive skills tests for people in their 40s and 50s, allows us to study this question, and our results reveal that these long-term effects are important.

In addition, this paper complements the literature on radiation on human capital. The works of Almond et al. (2009) and Black et al. (2013), both based on detailed administrative data from Scandinavia, show that in-utero exposure to radiation lowers cognitive test scores at school-leaving age, and these effects even transmit to the next generation. Our paper shows that radiation can have equally strong effects when people are constantly exposed throughout their adult lives. A series of papers exploits the Chernobyl accident to provide evidence on the effects of radiation in Ukraine, the country where Chernobyl is located. Lehmann and Wadsworth (2011) find negative effects on self-reported health and labor market performance, Danzer and Danzer (2016) find negative effects on mental health, and provide evidence that radiation exposure affects people's preferences. Related to this literature, our contribution is twofold. First, we show that the Chernobyl-induced radiation significantly affected people's cognitive skills. Moreover, given that German is located more than 1000km West of Chernobyl, our findings suggest that nuclear power generation can have substantial external costs.

The remainder of the paper is structured as follows. In Section 2, we provide the background of the Chernobyl nuclear disaster and summarize the medical literature on the effect of radiation. In Section 4, we describe the dataset and show descriptive statistics. Section 5 explains the identification strategy and discusses potential threats to identification. In Section 6, we present the main results as well as robustness checks. In Section 6.3, we provide evidence on some channels before concluding in Section 8.

2 Historical background

We begin with a description of the historical background, namely the Chernobyl nuclear disaster in Ukraine and its impact on radioactivity levels in Germany. We further describe the information the German population received in the aftermath of the disaster, which is key to our identification strategy. Finally, we summarize the biological, medical and psychological literature, which provides theories of the effects of radiation on the human body, and how these can translate into cognitive test scores.

2.1 The Chernobyl disaster and its impact in Germany

The Chernobyl nuclear disaster 1986 The Chernobyl nuclear disaster in 1986 was one of the largest accidents in the history of nuclear energy. It occurred after a failed simulation of a power cut at a nuclear power plant in Chernobyl/Ukraine on April 26, 1986, which triggered an uncontrolled chain reaction and led to the explosion of the nuclear reactor. In the two weeks following the accident, several trillion Becquerel of radioactive matter were emitted from the reactor, stirred up into the atmosphere, and, through strong East winds, carried all over Europe.[1] The most affected countries were Belarus, Ukraine as well as the European part

[1] Becquerel (Bq) is a unit of radioactivity. One Bq defines the activity of radioactive material in which one nucleus decays per second. In the following, we use kilobecquerel (kBq). One kBq equals 1000Bq.

of Russia, although other regions, such as Scandinavia, the Balkans, Austria and Germany also received considerable amounts of nuclear fallout. The only other nuclear accident with comparable levels of fallout was the Fukushima disaster in Japan in 2011 (Yasunari et al., 2011).

Post-Chernobyl radiation in Germany The radioactive plume reached Germany three days after the disaster, on April 30, 1986. It first entered the country in the southeast and made its way northwest before disappearing over the North Sea on May 10. The trajectory of the plume is illustrated in Figure 1, which shows the air concentration of radioactive particles (radionuclides) in four measuring stations in different parts of Germany. The station Brotjacklriegl, a mountain in the southeast, close to the border with the Czech Republic and Austria, is located in the area that was first reached by the plume. A high air concentration of caesium-137 was registered on April 30, which faded after two days. The stations in Neuherberg — close to Munich, further to the northwest — and Offenbach — close to Frankfurt, in the center of the country — registered a high concentration around May 2/3, whereas in Norderney, an island in the North Sea, a marginally higher concentration was only measured on May 4.

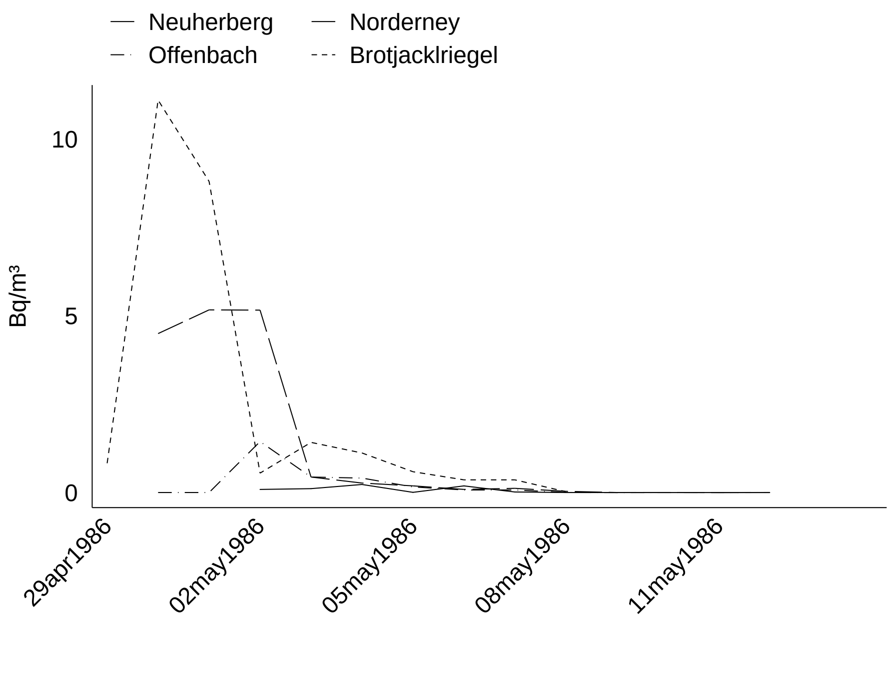


Figure 1: Air concentration of radioactive particles in 1986

Notes: This graph displays the air concentration of Cs137 measured after the arrival of the radioactive plume in four German measuring stations. These are located in different parts of the country: Brotjacklriegel (south-eastern border), Neuherberg (south-east), Offenbach (center) and Norderney (north-west). Source: Federal Office for Radiation Protection (Bundesamt für Strahlenschutz).

The deposition of the fallout varies considerably across regions, and depends on the amount of rainfall within a critical time window. Regions that experienced heavy rainfall while the radioactive plume was hanging above Europe received large amounts of fallout whereas regions without rainfall received very little. For affected regions, the nuclear fallout represented a long-lasting shock to radiation levels. While the air concentration of radioactive particles vanished after a few days, the ground deposition remains in the soil until today. The fallout consists of four main isotopes, namely caesium-137 (Cs137), caesium-134 (Cs134), strontium-90 (Sr90) and iodine-131 (I131), which have half-lives of up to 30 years.[2] Among the four isotopes, soil-bounded Cs137 is considered today the only relevant source of radiation in Germany that can be ascribed to the Chernobyl disaster (Hachenberger et al., 2017). In 2010, the first year in which we measure people's cognitive skills, more than half of the fallout was still in the ground, although, over time it has been washed out into deeper layers of soil, thereby reducing the external exposure of the population (Bunzl et al., 1995). However, exposure through ingestion is possible until today, as certain foods, in particular mushrooms and game, still exceed radiation limits in parts of South Germany.

Figure 2a displays the ground deposition of Cs137 as measured in May 1986. Because Cs137 rarely occurred in Germany before 1986, the displayed variation is almost completely induced by the Chernobyl fallout. The regions that received the highest level of fallout were Bavaria and Baden-Wuerttemberg in the south as well as parts of the former German Democratic Republic. Across Germany, the level of ground deposition ranges from 0.224 kBq/m^2 to 107 kBq/m^2, whereas soil is officially considered contaminated if the radioactivity exceeds 37 kBq/m^2 (UNSCEAR, 2000). Most Germans lived in areas with radiation levels below 20 kBq/m^2, although a non-negligible number of people lived in areas with levels much higher than that.[3]

The additional, Chernobyl-induced, effective dose of radiation is comparable to that from ordinary sources. In the first five years, the cumulative effective dose received by the average German person was 0.275 mSv, which is 13 times the dose of a chest x-ray, three times the dose of a round trip Frankfurt-New York City, or around 70% of the dose of a mammogram.[4] In Munich, one of the more heavily exposed cities, the additional effective dose over five years was 0.905 mSv, whereas in the most heavily exposed regions, the average person received an additional effective dose of 2.33 mSv (Bundesregierung, 1986-1991).[5]

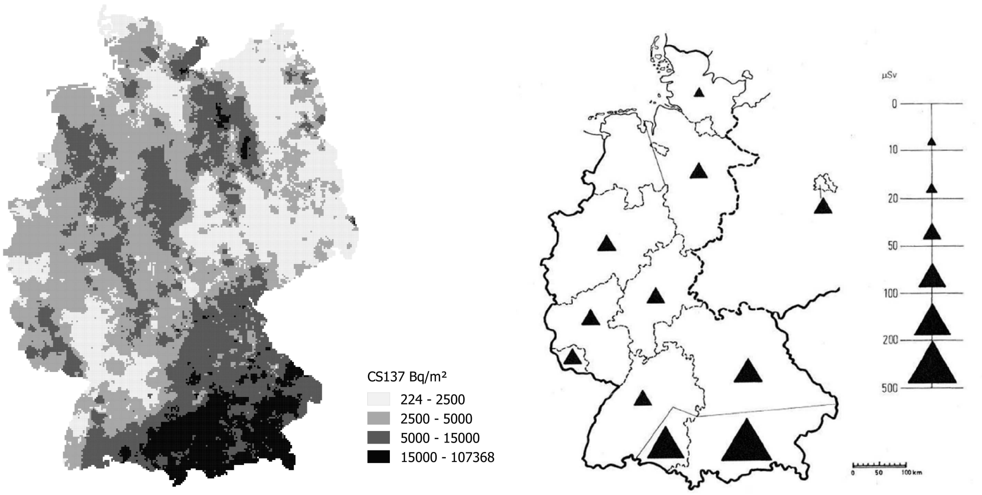
Information about the nuclear disaster in the German public The German public only learned about the nuclear accident several days after it occurred, and — in most parts of the country — after the radioactive rain had fallen. Indications of a nuclear accident were

[2]The half-lives of the four isotopes are eight days (I131), two years (Cs134), 28.8 years (Sr90), and 30.2 years (Cs137). We will use the abbreviations in parentheses further in the paper. These do not correspond to the abbreviations used in chemistry, which are ^{137}Cs, ^{134}Cs, ^{90}Sr and ^{131}I.

[3]See Figure 5b in Appendix B.

[4]For estimates of effective doses during medical procedures, see Mettler et al. (2008). A millisievert is the unit of effective dose. Effective dose is a measure of ionizing radiation, weighted for both the quality of radiation in question and the tissue response to radiation.

[5]The effective dose received during one x-ray is comparable in units to the effective dose received by the average person during a year as health effects seem to be unrelated to the length of low-dose exposure (Leuraud et al., 2015). It should be noted, however, that the average exposure published by Bundesregierung (1986-1991) is more uncertain and is based on assumptions about daily activities, diet, etc..



(a) Actual ground deposition

(b) Information released in 1986

Figure 2: Ground contamination in 1986

Notes: These graphs display (a) the ground deposition of Cs137 in Bq/m^2 and (b) the information about regional exposure in mSv that was released to the public in 1986. Source: Federal Office for Radiation Protection (Bundesamt für Strahlenschutz), , German-Swiss Association for Radiation Protection (Fachverband für Strahlenschutz e.V.)

first noticed in Sweden, where scientists measured abnormally high levels of radioactivity at the Forsmark nuclear power plant, despite the plant running as usual. The Soviet Union initially released no information about the accident, and its government only acknowledged the accident after the information from Sweden had spread. The German population was officially informed for the first time during the newscast "Tagesschau" on April 29, which reported about high levels of radioactive matter being emitted from an exploded nuclear power plant in Ukraine. In the same newscast, the Federal Minister of the Interior, Friedrich Zimmermann, stated that, due to the distance to Ukraine, there was no danger for the German population. However, two days later, after high radiation levels have been measured in several parts of the country, the government of the FRG introduced radiation limits on foods and warned the population of the consumption of dairy products, vegetables, mushrooms and game, which were potentially contaminated. In the following days, contaminated food was discarded and public swimming pools and playgrounds were temporarily closed. Despite these measures, the German government maintained its official communication that the increased radiation does not present a health hazard to the population. The information policy and the dispersion of information differed considerably between the Federal Republic of Germany (FRG) and the German Democratic Republic (GDR). In the GDR, no comparable measures were put in place. Quite the opposite, after the accident and the collapse of demand in the FRG, agricultural products intended for export to the FRG were supplied to the market in the GDR.

While the German population was generally informed about the radioactive fallout, they had little knowledge about the levels of fallout in particular areas. Figure 2b shows a map that was released by the Federal Office for Radiation Protection in 1986, which displays the average exposure in mSv in twelve large regions. A detailed map, such as the one shown in Figure 2a only became available five years later, in 1991. While there is plenty of anecdotal evidence that people changed some behaviors — diet, physical activity, time spent outside — in the immediate aftermath of the disaster, it appears that these changes were short-lived. Ortwin (1990), for example, shows that Germans' attitudes in favor of nuclear energy reverted to their pre-1986 levels one year after the accident.

3 Radiation and cognitive test scores: conceptual framework

To guide our empirical analysis, we build a simple conceptual framework that relates radiation exposure to cognitive skills through several causal channels. The framework is based on insights from medicine, radiobiology and psychology, which we briefly summarize.

3.1 Radiation, health and cognition: summary of the scientific literature

Sources of exposure Radiation has natural as well as artificial sources. Examples for natural sources are cosmic radiation, terrestrial radiation — emitted from radionuclides in the soil, such as uranium or potassium — or radon, a radioactive gas emitted from building materials. In addition, there are many artificial sources, such as nuclear power generation, nuclear weapons, or medical procedures such as x-rays, mammograms, or CT scans. Humans can be exposed to radiation in three ways, namely through the inhalation of radioactive particles, the ingestion of contaminated foods, as well as external exposure, whereby radiation affects the body if a person simply is present in a place with a given level of radioactivity in the environment. Exposure to radiation through air and ground can be directly assigned to a person's place of residence, and, thus, are highly correlated (Clark and Smith, 1988). Exposure through food, in contrast, may not necessarily result from contamination in the same locality as the food may have been produced elsewhere.

In the northern hemisphere, the average yearly exposure to natural radiation is 2.4 mSv, of which 52% is through inhalation, 12% through ingestion, and 36% through terrestrial and cosmic radiation (UNSCEAR, 2008). The degree of exposure differs between people and depends on their daily activities as well as their diet. For example, people who spend more time outdoors are more exposed to cosmic radiation than people who spend most of their time indoors; or people who are physically active and, therefore, breathe more, have a higher exposure through inhalation. The exposure to artificial radiation considerably differs between sources. An x-ray of the chest (0.02 mSv) or a return flight Frankfurt-New York (0.1 mSv) comprise low doses of radiation, while the dose received during a CT scan can be as high as 15 mSv. Euratom recommends that the annual effective dose of artificial radiation should not exceed 1 mSv, although this threshold excludes medical procedures.

Radiation exposure affects the human body through ionization, a process that damages the DNA and can lead to the dysfunction or death of cells. Radiation can directly damage the DNA in a cell when it collides and reacts with the DNA, or indirectly damage the DNA when it ionizes water molecules in the cell. Radiobiology theory posits that a marginal increase in radioactivity linearly increases the probability that a cell is hit by an electron. This linear relationship emerges because, during ionization, the release of electrons follows a random process, such that each cell has an equal likelihood of being hit. A marginal increase in radioactivity increases this likelihood and leads to a greater number of cells being hit. In addition, an infra-marginal increase in radioactivity increases the likelihood of the same cell being hit more than once. Cells that received the same amount of hits, in turn, are subject to the same types of damage and the same radiobiological processes (Brenner et al., 2003). A single hit will unlikely result in the death of a cell, but is sufficient to damage the cell's DNA, that is, the information content of a cell. The human organism has the capacity to repair damaged DNA, but if the DNA is not fully repaired, the cell may continue to regenerate and differentiate, thereby passing on the damaged DNA to future cell generations. This process can lead to mutations as well as the dysfunction of cells. The greater the number of affected cells and the longer the observation period is, the more likely it is that a critical mass of dysfunctional cells negatively affects health. The replacement and repair of damaged cells is prone to a stochastic error that increases with age (UNSCEAR, 1994), which is why the impact of radioactivity may be felt more by older than by younger people.

Health impacts A large body of literature in the sciences points to negative health effects of radiation exposure. The health effects can either be *deterministic*, whereby exposure to radiation almost inevitably affects a person's health, or *stochastic*, in which case radiation only affects the likelihood of developing a health condition. Deterministic effects only result from exposure to high doses of radiation, such as as those encountered by survivors of the Hiroshima nuclear bomb, or soldiers that were sent to the Chernobyl reactor to clear up the nuclear waste.[6] Exposure to a high dose — even for just a few minutes — induces reactions in the human body within days or weeks, such as damage to blood-forming organs, the stomach, the intestinal tract as well as the central nervous system. These reactions can lead to severe dysfunction of the human body, and are often fatal.

On the contrary, a low dose of radiation — defined as a short-term dose below 100 mSv — can only induce stochastic health effects. At low levels of radiation, an increase in the dose increases the probability that a person experiences health problems later in life, but it does not lead to the immediate dysfunction of organs (OECD, 2016). This is the case because low-dose radiation is not strong enough to kill cells, but can only damage their DNA (Stewart et al., 2012). Therefore, when exposed to a low dose of radiation, each cell only has a small likelihood of being hit in the first place, and once hit, there is a chance that it gets fully repaired by

[6] The ICRP sets the threshold for deterministic effects to 500mSv (Stewart et al., 2012).

the organism. A meta-analysis by Møller and Mousseau (2013) shows that variation in natural background radiation accounts for 2% of the variance in mutations in humans and animals.

The medical literature provides evidence for the existence of stochastic health effects. At-risk occupations, such as flight attendants or workers at nuclear power plants, who receive an additional dose between 1 mSv and 2.5 mSv per year, are shown to have a higher risk of cancer (UNSCEAR, 2008). de Gonzalez and Darby (2004) estimate the proportion of cancer — which is attributable to diagnostic X-rays — on up to 1.8% per year. To describe the dose-response relationship between low dose radiation exposure and health, the International Council on Radiation Protection (ICPR) and the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) recommend to use linear no-threshold models.[7]

The impact of radiation on the human body varies with age. Children tend to be more vulnerable than adults, although this is mainly the case for deterministic and less so for stochastic effects (UNSCEAR, 2014; Bromet and Havenaar, 2007). The difference in the impact on children has several biological reasons. Children have smaller bodies, resulting in a higher external exposure of the inner organs.[8] In addition, internal exposure through inhalation and ingestion is more detrimental because childrens' organs are closer together than those of adults, such that radionuclides concentrated in one organ are more likely to irradiate other organs. On the other hand, the tissue of a child is more resistant than that of an adult.

Radiation and cognitive functioning The effect of radiation on cognitive and neuro-developmental functioning is an active area of research in the sciences (OECD, 2016). Traditionally, the adult brain was considered resistant to radiation, as it was difficult to prove that brain cells regenerate (Deng et al., 2010). In the last two decades, however, research on humans and animals challenged this view. Cell regeneration has been found in the hippocampus, the part of the brain that governs several types of memory, in particular crystallized intelligence and learning (Squire, 2009; Supekar et al., 2013). Several studies show that a higher exposure to radiation slows down the regeneration of brain cells, which, in turn, can impair cognitive performance. Experimental research on animals finds reductions in cognitive performance if cell regeneration is reduced by radiation (Rola et al., 2004). Moreover, studies on humans find that exposure to low-dose radiation during medical treatments — which equal the exposure received from a CT scan — increases the risk of cognitive impairments several months to years after the treatment (Hall et al., 2004; Douw et al., 2009; Monje and Dietrich, 2012). The impact of radiation on cognitive functioning depends on a person's age. It is stronger at older ages because cell regeneration declines and the brain becomes more vulnerable to stochastic effects

[7] To date, however, the empirical evidence on the functional form is mixed. Findings by Little et al. (2009) and Leuraud et al. (2015) suggest a linear relationship between radiation and the occurrence of cancer, while studies by Tubiana et al. (2009) and Hendee and O'Connor (2012) raise doubts about the validity of these findings and the existence of such a relationship.

[8] The inner organs of an adult are more protected from radiation compared to those of a child because an adult's organ is surrounded by more body mass (skin, fat, bones, etc).

as a person gets older (Spalding et al., 2013).

3.2 Conceptual framework

The scientific literature highlights two broad channels through which radiation exposure affects cognitive skills, namely through its impact on brain cells as well as on the functioning of organs. To fix ideas, we summarize both channels in a test score production function, which we augment with people's behavioral responses to radiation exposure,

$$y = F[I(B), H(B), B]. \tag{1}$$

In Equation (1), a test score y is produced with three inputs: a person's intelligence (I), a person's general health condition (H), as well as any choices people make in response to radiation exposure, summarized by B. We mainly think of B as compensating behaviors, that is, people change their behavior to limit or counteract the impact of radiation on their life outcomes. There are many possible responses, for example investment in education, moving to a less contaminated area, changes in one's diet, or changes in one's exercise habits. We assume that these behaviors can have a direct effect on test scores, but they can also affect a person's intelligence or health, thus having an indirect effect on test scores.

Total differentiation of Equation (1) with respect to radiation R yields the proportional change of a test score in response to a change in exposure to radiation,

$$\frac{dy}{dR} = \underbrace{\frac{\partial F}{\partial I}\frac{\partial I}{\partial R} + \frac{\partial F}{\partial H}\frac{\partial H}{\partial R}}_{\substack{\text{direct effects} \\ \text{(cognition, health)}}} + \underbrace{\frac{\partial F}{\partial B}\frac{\partial B}{\partial R} + \frac{\partial F}{\partial I}\frac{\partial I}{\partial B}\frac{\partial B}{\partial R} + \frac{\partial F}{\partial H}\frac{\partial H}{\partial B}\frac{\partial B}{\partial B}}_{\text{behavioral responses}} . \tag{2}$$

Equation (2) provides guidance for the empirical analysis. First, it helps us to generate hypotheses about the sign of the overall effect. We expect the direct effects of radiation on cognition and health to be either negative or zero, but not positive. In each channel, the first term ($\partial F/\partial I$ and $\partial F/\partial H$) is most likely positive. For a constant level of health, we would expect a person with higher intelligence to have higher test scores. Likewise, for a constant level of intelligence, we would expect a healthier person to have higher average test scores. On the other hand, the second terms ($\partial I/\partial R$ and $\partial H/\partial R$) are either negative or zero, given that it is implausible that radiation exposure improves human cells and, therefore, enhances intelligence or a person's health condition. For the behavioral responses, we expect the sign to be positive or zero. Although it is theoretically possible that people undertake actions that exacerbate the exposure to radiation — for example, by eating more mushrooms in more contaminated areas — such responses are not very plausible. What seems more plausible is that people switch to a healthier lifestyle to compensate for the negative health effects of radiation, leading to a positive sign of the effect of behavioral responses on test scores. Given that some terms in Equation (2) are negative while others are positive, the sign of the total effect remains an empirical question. A negative total effect is possible if the direct effects are strong enough to outweigh the effect of behavioral responses. But the total effect could also be zero if the Chernobyl-induced radiation was not strong enough to have any effect, or it could be positive if the effect of the behavioral responses outweighs the direct effects.

Second, Equation (2) helps with the interpretation of the estimates. A regression that relates cognitive test scores with radiation exposure — assuming that radiation exposure is exogenous — allows us to identify the *total* effect of radiation exposure on test scores, which comprises direct effects on intelligence and health as well as compensating behaviors. If one was interested in quantifying the importance of a particular channel, this would require controlling for all other channels or finding a quasi-experimental design in which the remaining channels are plausibly absent. In our analysis, while we are not able to fully disentangle the direct and indirect channels, our data allow us to test whether some plausible behavioral channel have an influence by testing whether $\frac{\partial B}{\partial R} = 0$.

4 Data and Descriptive Analysis

For our analysis, we link rich individual-level survey data with geo-coded information on radiation in a person's municipality of residence in the month before the disaster. In this section, we describe the construction of the dataset as well as the measurement of cognitive skills, and present descriptive statistics. We limit the description of the dataset to the most important aspects. In Appendix A we provide more detailed information and perform a large number of balancing tests to ensure that the estimation results are not driven by sample selection.

4.1 The NEPS data

Our main data source is the National Educational Panel Study (NEPS), a rich representative dataset on educational trajectories in Germany, supervised and hosted by the Leibniz Institute for Educational Trajectories (LIfBi, Blossfeld et al. (2011)). NEPS consists of six starting cohort, ranging from newborns to adults, which have been followed up in multiple survey waves since 2010. In this study, we use the adult cohort of NEPS (Starting Cohort 6 — SC6), which samples respondents born between 1944 and 1986. To set up the NEPS SC6, LIfBi took over a representative survey named Working and Learning in a Changing World (ALWA) that had been conducted by the Institute for Employment Research (IAB) in 2007 with originally 10,404 respondents, covering the birth cohorts 1956 to 1986.

NEPS SC6 includes all respondents of ALWA who were willing to enter the panel and be surveyed every year (N=8,997). Among the people who agreed to be included in the first place, 6,572 actually participated in the survey.[9] A comparison of the ALWA subsample with

[9] Of the 2,425 respondents who did not participate despite agreeing, 68% were unwilling, while 32% could not

the German Microcensus shows that, by and large, the sample is representative of the German population, although people with higher education and older people are slightly over-represented, whereas migrants are under-represented. When NEPS began its data collection in 2009/10, the ALWA sample was augmented with a representative refresher sample of the birth cohorts 1956 to1986, as well as a representative sample of cohorts born between 1944 and 1955.

Besides including rich data on personal characteristics, employment histories, and educational attainment, the NEPS SC6 offers two features that are key to our analysis. First, the survey includes a series of standardized competence tests, allowing us to measure cognitive skills along various dimensions. Second, the ALWA subsample includes detailed information on respondents' residential histories. In the initial survey wave, respondents were asked to provide monthly spell data on their municipality of residence since their birth. This allows us to link personal characteristics and cognitive test scores measured in the 2010s with data on radiation levels in the person's municipality of residence in May 1986, the time of the Chernobyl disaster.

4.2 Estimation sample

In total, we can link the municipality of residence in May 1986 for 5,827 participants. For the remaining 745 participants, we could not link the municipality information due to missing municipality keys (419 obs.) or because they lived abroad in May 1986 (326 obs.). Observations with missing municipality keys include people born after April 1986. Therefore, our estimation sample only includes people born before the disaster.

Because the residential history is only available in the ALWA subsample, we will base our analysis on this sample. In Appendix A.1, we provide a detailed description of the sample design and the actions taken by the interviewers to minimize recall error when eliciting data on respondents' educational, residential and work history. To reduce classification error, we drop from the sample respondents who moved in May 1986 (34 obs.), because we cannot determine whether they moved before or after the radioactive plume reached Germany. We additionally drop all respondents who did not participate in the competence tests (1,265 obs.), as well as all participants for which information on personal characteristics is missing (105 obs.). Our final estimation sample comprises 4,423 observations. To address concerns about the representativeness of the estimation sample, we perform a series of balancing tests in Appendix D, which suggest that the missing information is unsystematic.

4.3 Cognitive skills tests

One of the core objectives of the NEPS SC6 was to collect data on the competencies of adults. The survey includes eight standardized tests aimed at measuring competencies in reading, maths, information and communication technologies (ICT), and sciences. All competence tests included

be contacted.

in the NEPS SC6 were modeled after well-established tests that are frequently used in psychology and related fields (Weinert et al., 2011). For our analysis, we use the following tests:

- *Mathematical competence*, measuring skills in algebra, geometry and probability;
- *scientific literacy*, testing basic understanding of scientific concepts;
- *listening comprehension*, measuring the ability to listen to details;
- *ICT literacy*, testing the problem-oriented use of ICT;
- *reading speed*, measuring the speed of reading;
- *perceptual speed*, measuring the speed of processing information;
- *reasoning*, measuring basic logical reasoning.

Appendix A.2 provides a detailed description of each test. In the empirical analysis, we use each test score as a separate outcome. To make the estimates comparable across outcomes, we standardize the test scores to mean zero and standard deviation one. Moreover, given that the test scores measure different aspects of cognitive skills, we compute a standardized cognitive skills index that allows us to estimate the overall effect of radiation on the latent factor cognitive skills.

When studying the effect of radiation on cognitive skills, it is important to distinguish between the effect on crystallized intelligence, defined as the ability to use learned knowledge and experience, and that on fluid intelligence, defined as the ability to solve new problems, use logic in new situations, and identify patterns (Cattell, 1987). In the human brain, radiation mainly affects cells in the hippocampus, the part of the brain that is mainly responsible for governing crystallized intelligence. To test whether crystallized intelligence is more affected than fluid intelligence, we construct two sub-indices, each based on several tests. The index for crystallized intelligence comprises the standardized test scores of mathematical competence, reading, competence, ICT literacy, scientific literacy and listening comprehension. The index for fluid intelligence comprises reading comprehension, perceptual speed and reasoning.

4.4 Municipality-level data

Data on ground deposition Our main regressor of interest is the ground deposition of Cs137 in kBq/m^2, which we use as proxy for Chernobyl-induced radiation in Germany. The regional concentration of Cs137 is highly correlated with other Chernobyl-induced sources of radiation such as I131 or Sr90 (Hou et al., 2003), although compared to other radionuclides it is easier to measure and, due to its long half-life, is mainly responsible for long-run exposure of the population (International Atomic Energy Agency, 2006). The Federal Office for Radiation Protection (Bundesamt für Strahlenschutz, BfS) provided us with geo-coded data for the soil surface contamination in Germany at a 3×3km-grid-cell level. The data were compiled by the BfS based on 3,474 measurement points, at which the ground deposition of Cs137 was measured between 1986 and 1989 and decay corrected to May 1986. This density of measuring points makes Germany one of the countries with the most detailed data on radiation. Based on these measures, the BfS calculated the ground deposition for each grid cell as the distance-weighted average of the four nearest measurement points.[10]

Linkage between individual and regional data In an ideal scenario, we would measure for each person in the survey the *actual* individual radiation exposure over 25 years. In absence of such data, the second-best option is to link the survey data with information on a person's *potential* exposure. We do so by linking the NEPS with information on the ground deposition of Cs137 in the centroid of a person's municipality of residence in May 1986.[11] Compared to the first-best measure, our measure of potential exposure has two important limitations. First, the exact place of residence within a municipality is unknown. Assigning the person the radiation at the municipality level inevitably introduces measurement error in the regressor of interest because exposure is not uniform within a municipality. Second, we can only measure a person's *potential* exposure, while the actual dose may differ depending on people's lifestyles. We undertake several steps to assess the importance of these data constraints. We link the radiation data in several ways — with respect to the population center of a municipality, the population mode, or by only using the nearest measuring point — and show that the results are robust to the linkage procedure. In addition, to address measurement error, we carry out an instrumental variable estimation whereby we exploit idiosyncratic rainfall in the critical time period after the nuclear accident. Finally, although we can only measure potential exposure, we show that this measure is useful to provide an intent-to-treat interpretation.

Additional data We supplement our dataset with municipality-level data on precipitation, altitude, and population density. The precipitation data is collected on a daily basis by 544 stations across the country.[12] To link the precipitation data with the survey data, we compute for each municipality centroid the inverse-distance-weighted average rainfall based on the four closest measuring stations. To measure the average altitude of a municipality — an important determinant for rainfall as well as radiation — we use data provided by the Service Center of the Federal Government for Geo-Information and Geodesy. From the same source, we also obtained data on population size for all municipalities in Germany. The earliest available geo-

[10]See Figure 8a in Appendix F for the location of measurement points in West and East Germany.

[11]The German Federal Agency for Cartography and Geodesy (BKG) provided us with a list of all municipalities according to the definition as of 2013, their official municipality keys, as well as the geo-codes of the municipality centroids. Due to confidentiality issues, NEPS does not release the municipality keys to its users, but the LIfBI offers to merge data at the municipality level. We are very grateful for this service.

[12]The data are provided by the German Meteorological Service. See Figure 8 for the location of the measuring points.

coded data on the population size of municipalities is from 1997.[13] In line with LifBi's data protection policy, the municipality information has been merged by the research data center. To ensure that municipalities cannot be identified based on this information, we rounded the environmental data.[14]

4.5 Descriptive statistics

Table 1 displays the descriptive statistics of the main variables used in the regression. The information on personal characteristics, own and parental educational attainment reflects characteristics that were determined before 1986. Our sample only includes people born before the Chernobyl disaster. In 1986, the average person in the sample was 19 years old, with ages ranging from zero to 30 years. 37% of the sample, predominantly the older cohorts, were employed at the time, while another 42% attended an educational institution and 1% was unemployed. The share of people that lived in the GDR represents 19% of the sample. Only 2% of the sample are non-native speaker.

The German secondary school system distinguishes between three tracks, namely lower secondary school (*Hauptschule*, graduation after 9 years of schooling), intermediate secondary school (*Realschule*, 10 years), and upper secondary school (*Gymnasium*, 12 or 13 years). People with an upper secondary school degree can pursue a tertiary education, whereas people with lower degrees typically enter vocational training after graduating. Among those persons that were no longer in education in April 1986 around 4% had a lower secondary or secondary, while 29% and 13%, respectively, had an upper secondary or tertiary degree. An the other hand, 41% were still in education, most of whom received no degree (31%). However, 10% already passed lower secondary or secondary education and 1% passed upper secondary education. Nobody was in teartiary education.

The dataset also includes information on the highest school degree of the respondents' parents. The means reflect the seminal changes in the German education system, whereby the generations born in the 1950s and before had much lower educational attainment than their children. Over half of all respondents have parents with no more than 9 years of schooling.

The fourth set of statistics in Panel A refer to the cognitive test scores. Two features are important to note here. First, each test has a different metric, resulting in differences in means and standard deviations. Without a standardization, the estimates will be difficult to interpret. Second, the number of observations differs between tests, which is due to design features of NEPS explained in Section 4.3 as well as Appendix A.

Panel B displays the municipality-level characteristics. The statistics were computed across individual observations in the estimation sample. The ground deposition of Cs137, amounts to

[13]NEPS uses the official municipality codes of 2013. For municipalities that have been merged between 1997 and 2013, we use the sum of the merged municipalities.

[14]Cs137 rounded to the nearest 100, altitude rounded to the nearest 50, population rounded to the nearest 5000, precipitation rounded to the first decimal place

$5.17 kBq/m^2$. The standard deviation, which is larger than the mean, points to a significant variation in ground deposition across Germany.[15] The level of precipitation represents the average rainfall in May in the five years preceding the Chernobyl disaster, i.e. 1981-1985. The average person lived in 1986 in a medium-sized municipality with 282,000 inhabitants, although municipality sizes vary between 5,000 and over 3 million.

[15]See Appendix B for an illustration of the distribution of the ground deposition across municipalities.

Table 1: Descriptive statistics of the main variables

	Mean	(SD)	min	max	N
A. Individual-level data					
Personal characteristics					
Age in 1986	19.05	8.20	0.00	30.43	4423
Female	0.51	0.50	0.00	1.00	4423
Native speaker	0.98	0.15	0.00	1.00	4423
GDR	0.19	0.39	0.00	1.00	4423
Unemployed in April 1986	0.01	0.12	0.00	1.00	4423
Employed in April 1986	0.37	0.48	0.00	1.00	4423
Educational attainment in April 1986					
Not of school age yet (less than 7 years old)	0.12	0.33	0.00	1.00	4423
No degree, lower secondary, secondary	0.04	0.19	0.00	1.00	4423
Upper secondary	0.29	0.45	0.00	1.00	4423
Tertiary	0.13	0.33	0.00	1.00	4423
In school or college education,	0.42	0.49	0.00	1.00	4423
no degree	0.31	0.47	0.00	1.00	4423
already attained lower secondary, secondary	0.10	0.30	0.00	1.00	4423
already attained upper secondary	0.01	0.08	0.00	1.00	4423
already attained tertiary	0.00	0.00	0.00	1.00	4423
Highest parental education					
No degree, lower secondary	0.52	0.50	0.00	1.00	4423
Secondary	0.27	0.44	0.00	1.00	4423
Upper secondary	0.21	0.41	0.00	1.00	4423
Test Scores					
Math	11.32	4.75	0.00	21.00	2644
Reading	27.07	7.45	0.00	39.00	2659
Reading Speed	38.19	8.33	0.00	51.00	3602
Scientific literacy	18.05	5.34	0.00	28.00	3275
ICT	41.19	13.62	0.00	66.00	3298
Reasoning	8.94	2.37	0.00	12.00	3157
Listening comprehansion	75.82	7.97	0.00	89.00	3160
Perceptual Speed	34.69	8.06	0.00	82.00	3158
B. Municipality-level data					
Caesium137 kBq/m^2 (01. May 1986)	5.17	5.86	0.50	62.10	4423
Precipitation mm/m^2 (yearly average, 1981-1985)	3.09	0.84	1.30	8.00	4423
Altitude in meter	201.58	176.65	0.00	850.00	4423
Minimum altitude in meter in county	138.59	139.63	-1.00	660.00	4423
Population/1000	282.46	677.53	5.00	3420.00	4423

Notes: This table displays the descriptive statistics for the most important variables. The number of observations varies between tests due to the survey design. See Appendix A for a comprehensive description of the testing procedure. The data underlying the statistics in Panel B are measured at the municipality level, although the statistics themselves are computed at the individual level.

5 Empirical Strategy

In this section, we present the empirical model and the identifying assumption that is necessary to interpret the estimated relationship as causal. We further discuss two important threats to identification, namely anticipation and residential sorting, and show balancing tests that compare the observable characteristics of people with high and low exposure to radiation. Finally, we discuss challenges to statistical inference due to clustering and multiple hypothesis testing.

5.1 Empirical model

Our aim is to estimate the impact of exposure to caesium-137 on cognitive skills. To this end, we estimate the following empirical model.

$$y_{ims} = \alpha + \beta \; CS137_{ms} + \boldsymbol{X'_{im}}\gamma + \delta_s + \varepsilon_{ims}. \tag{3}$$

The cognitive test score y_{ims} of person i who resided in municipality m in State s in April 1986 is regressed on the ground deposition of Cs137 in the same municipality. The vector $\boldsymbol{X_{im}}$ controls for pre-treatment characteristics of individuals and municipalities. At the individual level, it includes controls for gender, age, country of origin, parental education, education in 1986 and employment status in 1986. It also includes municipality characteristics, namely the average level of rainfall, altitude, and population size. In some specifications, we will also control for fixed effects at the level of federal states, δ_s. The error term ε_{ims} summarizes all determinants of cognitive test scores not captured by the regressors.

In line with the conceptual framework in Section 3, our coefficient of interest, β, measures the total effect of exposure to radiation in April 1986 on cognitive test scores. However, because the ground deposition measures the potential rather than the actual exposure of a person, the regression in Equation (5) represents the reduced form of the total effect.

5.2 Identification

The parameter β can only be interpreted as causal if the ground deposition of Cs137 is uncorrelated with any determinants of test scores that are not controlled for in Equation (5), i.e. $E(\varepsilon_{im} \times CS137_{im} | \boldsymbol{X_{im}}, \delta_s) = 0$. We believe that this assumption is plausible because the German population could neither foresee the nuclear disaster nor the rainfall patterns in the critical time window after the disaster. Nonetheless, there are several potential confounding factors that could violate the identifying assumption. We discuss these factors below and outline how we corroborate the validity of the identifying assumption.

Anticipation One threat to identification is anticipation. If people knew that their area would be contaminated, they could move to a different area to avoid exposure. Anticipation would confound our estimates because the location choice would determine the exposure but, at the same time, may directly impact cognitive test scores, thereby violating the identifying assumption. In the context of the Chernobyl disaster, however, anticipatory behavior is unlikely, because Germans could neither foresee that the nuclear disaster would happen, nor that the plume would be carried to Germany. Moreover, because the ground deposition depended on rainfall in a critical time window of five days in late April and early May 1986, it is not plausible that people could anticipate which particular areas would be contaminated.

Importantly, anticipation — an ex-ante change in behavior in expectation of contamination — is not to be confused with ex-post avoidance behavior. It is possible that people responded to high exposure with changes in their lifestyle, for example by changing their diet or exercise habits. Although such behavioral responses affect the effective exposure of people, they do not invalidate the identifying assumption but rather represent a channel through which radiation affects test scores.

Residential sorting A potentially more severe threat to identification is residential sorting. Areas differ with respect to the amenities and job opportunities they provide, which is why some people prefer living in urban while others prefer living in rural areas. The determinants of this sorting process may be correlated with the ground deposition in an area, which can confound the estimation even if people do not deliberately move to a place to avoid radiation. An example for such a determinant is rainfall. Suppose that areas with less rainfall may be more attractive to skilled than to unskilled people. On the other hand, areas with less rainfall have a lower likelihood of receiving nuclear fallout. Consequently, the sorting of skilled people into places with low rainfall would lead to a spurious correlation between fallout and cognitive skills.

Balancing tests in fact point to residential sorting. In Columns (1)-(3) of Table 2, we compare the pre-treatment characteristics of people who lived in 1986 in municipalities with below- and above-median ground deposition of Cs137. While the sample is balanced on some characteristics, there is evident sorting based on skills. People with a low education as well as those whose parents have a low education were more likely to live in areas that received a higher level of fallout. Panel B provides some potential reasons for this sorting pattern. Municipalities with above-median levels of fallout tend to have a higher altitude, tend to be less populated, and have more rainfall. In other words, less skilled Germans tend to live in more rural areas, and rural areas received a greater level of nuclear fallout after the Chernobyl disaster due to their altitude and rainfall levels.

In Columns (4)-(9), we test whether the sample is balanced conditional on controls, which can inform the appropriate set of control variables in our regression. The results in Columns (4)-(6) suggest that controlling for altitude, rainfall and population size cannot fully eliminate residential sorting. In Columns (7)-(9), we additionally control for state fixed effects, restricting the comparisons to within states. Conditional on these controls, the sample is balanced on all observable characteristics. Therefore, our preferred specification will include controls for municipality-level characteristics as well as state fixed effects.

Further threats to identification. Besides anticipation and residential sorting, there are at least three additional challenges to identification. First, while Table 2 suggests that the sample is balanced on observable characteristics once we control for municipality characteristics and state fixed effects, some unobserved differences between people living in high- and low-exposure areas may remain. A second challenge is selective attrition. Radiation can increase the risk of dying from cancer, potentially leading to a selected estimation sample. Likewise, not all respondents took part in all cognitive skills tests, and this non-participation is potentially systematic. Third, the linkage of radiation data with individual-level survey data introduces measurement error, because we only know the potential exposure in the person's municipality of residence, but neither the ground deposition in the exact location of residence nor the person's actual exposure.

We address these challenges using several methods, namely balancing tests, selection on observables, instrumental variables, as well as a series of robustness checks. We discuss the implications of these tests along with the main estimation results in the next section, and provide further details in the appendix. However, for the interpretation of the results to follow, we work with the maintained assumption that, conditional on state fixed effects and controls for municipality characteristics, the ground deposition of Cs137 is uncorrelated with any personal determinants of cognitive test scores.

5.3 Inference

Statistical inference about our estimate for β is challenging for two reasons, namely cross-sectional dependence of error terms and multiple hypothesis testing. The empirical model in Equation 5 is a regression of an individual variable on a group variable, such that the regressors are perfectly correlated between respondents living in the same municipality. Moreover, the unobserved determinants of cognitive test scores, ε_{ims}, can be correlated across individuals, for example if the instruction in schools differs across districts, leading to differences in test scores. To account for such cross-sectional dependence of error terms, we cluster the standard errors at the county level. Counties are one geographic level above municipalities, with the exception several large cities, where both units coincide. By clustering at the county level, we account for correlations within counties, but implicitly assume that the error terms are not correlated across counties within a state. Clustering at the state-level, in turn, would be complicated because Germany only has 16 federal states, such that the number of clusters would be too few to provide reliable estimates for cluster-adjusted standard errors. To assess our inference, we perform a non-parametric permutation test whereby repeatedly randomly assign the ground deposition of Cs137 within states.

A second challenge is multiple hypothesis testing. Our main analysis uses eight test scores as outcome variables. Because each test score measures a different dimension of the latent factor cognitive skills, the test scores are likely correlated. The correlation of outcome variables in turn increases the probability of obtaining at least one statistically significant result. To account for multiple hypothesis testing, we perform two separate analyses, one in which we adjust the standard errors of each estimate to account for mechanical correlations between outcomes (Benjamini and Hochberg, 1995), and one in which we summarize all results in a weighted index, such that we only need to test one hypothesis (OB́rien, 1984; Anderson, 2008).

Table 2: Sample description: Balancing table

	Raw data			Control municipality charac.			State FE, Municipality charac.		
	Mean		Diff (2)-(1)	Mean		Diff (5)-(4)	Mean		Diff (8)-(7)
	Below	Above		Below	Above				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Individual characteristics									
Age in 1986	18.950	19.152	0.202	-0.247	0.232	0.479*	-0.152	0.143	0.295
	(0.179)	(0.170)	(0.247)	(0.178)	(0.169)	(0.246)	(0.178)	(0.169)	(0.245)
Female	0.517	0.499	-0.017	0.007	-0.006	-0.013	0.003	-0.003	-0.006
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
Native speaker	0.974	0.980	0.006	-0.001	0.001	0.003	-0.001	0.001	0.001
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Employed in April 1986	0.375	0.374	-0.001	0.000	-0.000	-0.001	-0.005	0.004	0.009
	(0.010)	(0.010)	(0.015)	(0.010)	(0.010)	(0.015)	(0.010)	(0.010)	(0.014)
Unemployed in April 1986	0.016	0.014	-0.002	-0.000	0.000	0.001	0.002	-0.002	-0.004
	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)
If employed: Qualified or highly qualified job before May 1986	0.321	0.332	0.011	0.000	-0.000	-0.001	0.007	-0.007	-0.014
	(0.018)	(0.018)	(0.026)	(0.018)	(0.018)	(0.026)	(0.018)	(0.018)	(0.026)
Children before 1986	0.157	0.156	-0.001	-0.004	0.004	0.008	-0.013	0.012	0.024
	(0.014)	(0.014)	(0.020)	(0.014)	(0.014)	(0.019)	(0.014)	(0.014)	(0.019)
Older siblings	0.520	0.546	0.026*	-0.007	0.006	0.013	-0.008	0.007	0.015
	(0.011)	(0.011)	(0.016)	(0.011)	(0.011)	(0.016)	(0.011)	(0.011)	(0.015)
Smoke before 1986	0.402	0.425	0.023	-0.018	0.017	0.035**	-0.010	0.009	0.019
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
Educational attainment in April 1986									
In school or college education	0.418	0.424	0.005	-0.003	0.003	0.006	-0.001	0.001	0.002
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
Not of school age yet (less than 6 years old)	0.127	0.118	-0.009	0.005	-0.005	-0.010	0.003	-0.003	-0.005
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
Lower secondary and secondary	0.032	0.045	0.013**	-0.005	0.004	0.009	-0.001	0.001	0.002
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
Upper secondary	0.285	0.297	0.012	-0.004	0.003	0.007	0.000	-0.000	-0.000
	(0.010)	(0.010)	(0.014)	(0.010)	(0.010)	(0.014)	(0.010)	(0.010)	(0.014)
Tertiary	0.138	0.116	-0.022**	0.006	-0.006	-0.012	-0.000	0.000	0.001
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
Highest parental education									
Lower secondary education	0.466	0.576	0.110***	-0.031	0.029	0.060**	-0.010	0.010	0.020
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
Secondary education	0.288	0.248	-0.041***	0.007	-0.006	-0.013	0.000	-0.000	-0.000
	(0.010)	(0.009)	(0.013)	(0.010)	(0.009)	(0.013)	(0.010)	(0.009)	(0.013)
Upper secondary	0.246	0.176	-0.070***	0.024	-0.023	-0.047**	0.010	-0.010	-0.020
	(0.009)	(0.008)	(0.012)	(0.009)	(0.008)	(0.012)	(0.009)	(0.008)	(0.012)
B. Municipality characteristics									
Altitude in meter	141.312	257.954	116.641***						
	(2.861)	(4.055)	(4.963)						
Minimum altitude in meter in county	88.203	185.745	97.542***						
	(1.710)	(3.462)	(3.861)						
Population	439.750	134.888	-304.862***						
	(19.482)	(6.014)	(20.389)						
Precipitation in mm/m^2	2.911	3.265	0.353***						
	(0.021)	(0.014)	(0.025)						
GDR	0.313	0.078	-0.235***						
	(0.010)	(0.006)	(0.011)						
N		2141	2282						

Notes: This table displays the pre-treatment characteristics of individuals (Panel A) and municipalities (Panel B) in areas with above- and below-median ground deposition of Cs137. Columns (1) and (2) display the raw means and standard deviations, whereas Columns (4) and (5) as well as (7) and (8) display the residual means and standard deviations after conditioning on municipality characteristics and state fixed effects. In Columns (3), (6), and (9), we perform t-tests for equality in means. Standard errors of the test statistics are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

6 Exposure to radiation and human capital: results

In this section, we present the estimation results for the impact of radiation exposure on cognitive test scores. After showing and discussing the main results, we explore the non-linear dose-response relationship as well as heterogeneous effects across demographic groups. We further provide evidence that the observed results cannot be explained by behavioral changes such as internal migration or investment in education.

6.1 Main results

Table 3 reports the main estimation results for the impact of exposure to radiation on cognitive test scores. Each entry represents an estimate for β from a separate regression of the outcomes listed on the left on the ground deposition of Cs137 in kBq/m^2 and the controls listed at the bottom. The table comprises two sets of results; in Columns (1)-(4) the outcomes are the raw test scores, whereas in Columns (5)-(8), the outcomes have been standardized to mean zero and standard deviation one. In each set, we begin with a bivariate regression and gradually add control variables. Given the results from the balancing test presented in Section 5, our preferred specification includes controls for municipality characteristics as well as state fixed effects.

Once we control for individual characteristics, we consistently find negative coefficients, suggesting that exposure to radiation over 25 years has a negative impact on cognitive skills. However, the effect sizes differ between outcomes. Based on the results in Column (4), an increase in ground deposition by one standard deviation (5.86 kBq/m^2) decreases math scores by 0.3 points, which is 2.6% of the mean. For reading, the corresponding effect is -0.61 points, or 2.2% of the mean reading score. The effect sizes for the other outcomes range between -0.1% and -1.05% of the mean test scores. Four out of eight estimates are statistically significant at the 5%-level or lower.

To make the estimates comparable between test scores, in Columns (5)-(8), we report the estimated effects on standardized outcomes. Each coefficient denotes the impact of an increase in ground deposition by 1 kBq/m^2 measured in terms of standard deviations of the outcome. For example, a coefficient of -0.01 means that an increase in ground deposition in 1986 by one unit leads to a decrease in the test score by 1% of a standard deviation. The point estimates suggest that radiation has the largest impact on reading scores, followed by math, listening comprehension, and reading speed, while it has smaller impacts on ICT skills, scientific literacy, perceptual speed, and reasoning.

In Panel B, we additionally report the estimated impacts of radiation on a standardized cognitive skills index, as well as indices for skills relying on crystallized and fluid intelligence. For all three indices, the effects are large and statistically significant. For a one-unit increase in ground deposition, the standardized index decreases by 0.8% of a standard deviation. The effect on skills based on crystallized intelligence is larger (-0.9% of a standard deviation) than the one on skills based on fluid intelligence (-0.6% of a standard deviation).

Overall, these results suggest that continuous exposure to radiation over 25 years decreased people's cognitive skills, and that these negative estimates are statistically significantly different from zero. To evaluate the economic significance of our estimates, it is useful to consider the estimated impact of a one-standard-deviation increase in ground deposition in 1986. This decreases reading scores 25 years later by 8.2%, math scores by 6.5%, listening comprehension scores by 5.3% and reading speed by 4.7%, and ICT scores by 2.9% of the respective standard deviation. The impact on the overall cognitive skills index is -4.7% of a standard deviation. These results point to a high economic significance of the impact of radiation on cognitive skills.

6.2 Non-linear and heterogeneous effects

Non-linear dose-response relationship. In Table 4, we analyze if there is a non-linear dose-response relationship between radiation exposure and cognitive test scores. In each regression, the outcome is the cognitive skills index. For comparison, Column (1) reproduces the linear estimate reported in Column (8) of Table 3.

The estimates in Columns (2) and (4) provide little evidence in favor of a non-linear relationship. In Column (2), we impose a quadratic relationship, but find no significant coefficient for the quadratic term. In Column (4), we estimate a spline regression by interacting the ground deposition with a binary indicator that equals unity if a person lived in 1986 in an area with above-median ground deposition. While the point estimate is larger for people living in areas with above-median ground deposition, the coefficient is statistically insignificant, such that a linear relationship cannot be rejected. In Column (3), we impose a log-linear relationship, for which we find a large and statistically significant coefficient. For a one-standard-deviation increase in the log ground deposition (sd=0.72), we find a decrease in cognitive test scores by 5.6% of a standard deviation, which is similar to the estimate from the linear level-level model in Column (1).

While the level-level model in Column (1) and the level-log model in Column (3) have a similar fit, a level-level model makes more sense from a scientific standpoint. Radiobiology provides theories of a linear relationship between radiation exposure and the likelihood of a cell being damaged that have been verified in a series of experiments (Brenner et al., 2003). To the extent that our estimate is explained by the damage of brain cells or other cells in the body, it is plausible that radiation linearly affects test scores.

Heterogeneous effects In Table 5, we explore whether the impact of radiation exposure on cognitive skills differs between demographic groups. For each set of groups, we estimate full interaction models that interact the ground deposition of Cs137 with mutually exclusive dummies for each group but without including the dummies themselves. For example, in Column (1), we interact the ground deposition with a dummy for male and a dummy for female, which provides

Table 3: OLS results: the effect of radiation on cognitive skills

	Non standardized outcomes				Standardized outcomes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Individual test scores								
Math	0.014	-0.005	-0.055***	-0.052***	0.003	-0.001	-0.012***	-0.011***
mean =11.324	(0.018)	(0.013)	(0.016)	(0.015)	(0.004)	(0.003)	(0.003)	(0.003)
Reading	-0.012	-0.036	-0.102***	-0.104***	-0.002	-0.005	-0.014***	-0.014***
mean =27.070	(0.042)	(0.034)	(0.034)	(0.037)	(0.006)	(0.005)	(0.005)	(0.005)
Listening comprehension	-0.027	-0.050*	-0.070**	-0.073**	-0.003	-0.006*	-0.009**	-0.009**
mean =75.822	(0.030)	(0.026)	(0.029)	(0.032)	(0.004)	(0.003)	(0.004)	(0.004)
ICT	0.002	-0.039	-0.067	-0.075	0.000	-0.003	-0.005	-0.006
mean =41.186	(0.030)	(0.033)	(0.044)	(0.046)	(0.002)	(0.002)	(0.003)	(0.003)
Scientific literacy	0.004	0.000	-0.022	-0.021	0.001	0.000	-0.004	-0.004
mean =18.045	(0.015)	(0.013)	(0.018)	(0.017)	(0.003)	(0.003)	(0.003)	(0.003)
Reasoning	0.005	-0.006	-0.006	-0.002	0.002	-0.003	-0.002	-0.001
mean =8.941	(0.008)	(0.006)	(0.009)	(0.010)	(0.003)	(0.003)	(0.004)	(0.004)
Reading speed	-0.010	-0.047**	-0.089***	-0.067**	-0.001	-0.006**	-0.011***	-0.008**
mean =38.185	(0.028)	(0.020)	(0.029)	(0.030)	(0.003)	(0.002)	(0.004)	(0.004)
Perceptual speed	0.022	0.000	-0.032	-0.032	0.003	0.000	-0.004	-0.004
mean =34.691	(0.023)	(0.016)	(0.022)	(0.025)	(0.003)	(0.002)	(0.003)	(0.003)
B. Indices								
Cognitive skill index					0.000	-0.003	-0.008***	-0.008***
					(0.003)	(0.002)	(0.003)	(0.003)
Crystallized intelligence index					0.000	-0.002	-0.008***	-0.009***
					(0.003)	(0.003)	(0.003)	(0.003)
Fluid intelligence index					0.001	-0.003	-0.007**	-0.006*
					(0.003)	(0.002)	(0.003)	(0.003)
Controls:								
Individual characteristics	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Municipality characteristics	No	No	Yes	Yes	No	No	Yes	Yes
State FE	No	No	No	Yes	No	No	No	Yes

Notes: This table displays the main estimation results. Each coefficient is the result of a separate regression of the outcomes listed on the left on the ground deposition of Cs137 in kBq/m^2, controlling for the variables indicated below. Columns (1)-(4) report the results on the raw test scores; in Columns (5)-(8), the test scores have been standardized. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table 4: Non-linear effects

	(1)	(2)	(3)	(4)
CS137 kBq/m^2	-0.008***	-0.022**		-0.028
	(0.003)	(0.011)		(0.033)
CS137 kBq/m^2 $\times$ CS137 kBq/m^2		0.004		
		(0.009)		
ln(CS137 Bq/m^2)			-0.080***	
			(0.029)	
CS137 kBq/m^2 $\times$ above median				-0.110
				(0.089)
Controls:				
Individual characteristics	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Observations	4423	4423	4423	4423
Adj R^2	0.22	0.22	0.22	0.22

Notes: This table displays the estimates from OLS regressions of the standardized cognitive skill index on several functional forms of the ground deposition of Cs137 as well as the control variables listed at the bottom. See Section 5 for a detailed list of control variables. In Column (4), the ground deposition of Cs137 is interacted with an indicator that equals unity if a person lived in 1986 in an area with an above-median ground deposition. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

with separate estimates for both groups.[16] In all regressions, we control for individual and municipality characteristics as well as state fixed effects.

In Column (1), we find no difference in estimates between men and women. Despite potential differences between both genders in daily routines, exercise habits, and diets, we find the same point estimates for both groups.

In Column (2), we consider differences between age groups. We split the sample into three groups of similar size based on the age in May 1986, and generate mutually exclusive binary indicators, which we interact with the ground deposition. From this exercise, an interesting pattern emerges. While we find large negative effects for people aged 10 years and older in 1986, we find no effect on people who were younger than 10 years. At first glance this result runs counter to the findings of a large literature that shows that environmental insults in the womb and during early childhood have large negative effects on later-life outcomes (Almond and Currie, 2011). Moreover, the result is at odds with the findings in Almond et al. (2009) and Black et al. (2013), who show that children exposed as foetuses to high levels of radiation during a critical period of pregnancy have worse life outcomes compared to similar children who were in the womb a few months before the beginning of the exposure. Moreover, the cohorts born in

[16] We choose this specification for the ease of interpretation. It should be noted that, despite the inclusion of mutually exclusive dummies, there is no problem with multicollinearity. This would only occur if we additionally included the indicators in the regression.

the first half of the 1980s were young children at the time of the disaster, such that the exposure began in a critical phase of the development of their bodies. One potential explanation for this seemingly puzzling result is that the biological effects of exposure to radiation — those on brain cells and vital organs — are more likely to manifest themselves in older age. But because the youngest cohort is only 25 years old when taking the cognitive skills tests, we cannot observe what their test scores would be when the same cohort is in their mid-50s. Another potential explanation is that parents with young kids in 1986 particularly tried to shield their children away, thereby reducing the exposure.

In Column (3), we test for differences with respect to socio-economic status by comparing the effects on people whose parents have an education below and above secondary school (*Realschule*). The effect among people with less educated parents is almost three times as large as the effect for people with highly educated parents. There are many possible explanations for this difference. People of lower socio-economic status may have a greater exposure if they are more likely to work physically or through differences in their lifestyle. They may also have less knowledge or be less receptive to information about the negative consequences of radiation, such that they engage less in avoidance behavior.

In Column (4), we assess if the effects differ between people who, in 1986, lived in the GDR versus West Germany. Unlike in West Germany, the population in the GDR received little information about the disaster and its likely consequences, and was even encouraged to consume foods that were potentially contaminated. Given these differences, it is not surprising that the estimated effect in the GDR — although not statistically significant — is more than twice as large as the one for West Germany.

Finally, in Column (5), we test whether the effect differs depending on a person's health behavior before the disaster. NEPS SC6 includes information whether a person smokes and the year in which he or she started. Based on this information, we construct a binary indicator whether the person smoked in 1986. Given that we control for age and education in the regression, we ensure that the differences are not driven by these factors. We find a slightly larger estimate for people who smoked before 1986, although the difference between both coefficients is statistically insignificant.

6.3 Evidence on behavioral responses

Our main estimates, presented in Table 3 have to be interpreted as reduced-form results. The fallout in an area in 1986 affects a person's radiation exposure over 25 years, which, in turn affects cognitive test scores through its impact on cells as well as people's behavioral responses. In Table 6, we explore the importance of several behavioral responses. In our analysis we are constrained by the information available in our dataset. While NEPS SC6 has rich information on internal migration, employment, and education, we are not able to study several other behavioral responses, for example changes in health behaviors, diets, or exercise habits.

Table 5: Heterogenous effects

	(1)	(2)	(3)	(4)	(5)
CS137 kBq/m^2 × male	-0.008*				
	(0.004)				
CS137 kBq/m^2 × female	-0.008**				
	(0.004)				
CS137 kBq/m^2 × Age(0-10)		0.003			
		(0.005)			
CS137 kBq/m^2 × Age(10-20)		-0.018**			
		(0.007)			
CS137 kBq/m^2 × Age(>20)		-0.007**			
		(0.003)			
CS137 kBq/m^2 × Parent(above secondary education)			-0.004		
			(0.003)		
CS137 kBq/m^2 × Parent(below secondary education)			-0.010***		
			(0.003)		
CS137 kBq/m^2 × FRG				-0.008***	
				(0.003)	
CS137 kBq/m^2 × GDR				-0.016	
				(0.015)	
CS137 kBq/m^2 × Non-smoker before 1986					-0.007*
					(0.004)
CS137 kBq/m^2 × Smoke before 1986					-0.009**
					(0.004)
Controls:					
Individual characteristics	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Observations	4423	4423	4423	4423	4423
R^2	0.22	0.22	0.22	0.22	0.22

Notes: Each column reports the result from a regression of the standardized cognitive skills index on a full interaction between the ground deposition of Cs137 and mutually exclusive group indicators. In all regressions, we control for individual and municipality characteristics, as well as state fixed effects. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

In the first panel of Table 6, we study whether exposure to radiation triggered internal migration by using as outcome a binary indicator for whether, until a certain year, a person moved away from his or her municipality of residence in 1986. We regress this indicator on the fallout of Cs137 in 1986 as well as all other control variables and state fixed effects used in the base line regressions. The results provide evidence against internal migration as a behavioral response. This result is not surprising, given that a detailed map of ground contamination was only released to the general public five years after the disaster. Therefore, most people presumably were not aware of the contamination in their municipality of residence.

In the second panel, we consider labor supply as a behavioral response. As with migration, we find little evidence that people exposed to higher radiation levels were less likely to work. We find small and statistically insignificant effects on the number of months in employment. Likewise, we find little evidence that highly exposed people have a different likelihood of being employed at any point in time.

Finally, in the third panel, we estimate the impact on educational attainment, using as outcomes the years of education completed in a given year. For secondary and tertiary education, we find small and statistically insignificant negative effects, suggesting that formal education is not an important behavioral margin. However, we find a negative effect on the number of hours in continuing education — education people pursue while being employed. A one-standard-deviation increase in radiation decreases the average hours spent in 2010 in continuing education by 9 hours, which is 6.7% of the mean. Besides that, we find little evidence for the behavioral responses we are able to measure. In Appendix D, we additionally consider the impact of radiation on cohort-specific mortality, but find no evidence of any effect.

Table 6: Evidence on behavioral responses

	Coef.	(se)
Migration		
Until 1988	0.000	(0.001)
Until 1990	0.000	(0.001)
Until 1995	-0.002	(0.002)
Migration to lower exposure until 1995	0.000	(0.001)
Employment		
Month in employment between 1986 and 2010	-0.096	(0.380)
Employed in 2000	-0.001	(0.002)
Employed in 2005	-0.001	(0.002)
Employed in 2010	-0.000	(0.001)
Education		
Years in 1998	-0.001	(0.009)
Years in 1990	-0.007	(0.007)
Years in 1995	-0.009	(0.009)
Years in 2000	-0.008	(0.008)
Years in 2005	-0.006	(0.007)
Years in 2010	-0.008	(0.007)
Hours continuing education in 2010 (mean=134.5)	-1.571	(0.563)***
Controls:		
Individual characteristics	Yes	
Municipality characteristics	Yes	
State FE	Yes	

Notes: This table displays the results of separate regressions of the indicator variables listed on the left on the ground deposition of Cs137. In all regressions, we control for individual and municipality characteristics, as well as state fixed effects. For migration the outcome is an indicator that equals unity if, until a given year, a person moved away from his/her municipality of residence in 1986. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

7 Robustness checks

In this section, we summarize the results from a series of robustness checks that address challenges to our identification and inference. A detailed description of the results can be found in the appendix.

7.1 Addressing unobserved heterogeneity

The balancing tests in Section 5 show that conditional on state fixed effects and municipality-level controls for altitude, rainfall, and population size, the sample is balanced with respect to a large set of observable characteristics. Nonetheless, we are not able to observe all determinants of test scores. Because these unobserved determinants could be correlated with radiation exposure, it is possible that the omission of unobservables can bias the estimates.

7.1.1 Selection on observables

We first assess the potential influence of unobservable characteristics on our estimates. characteristics. Oster (2014) provides a method that quantifies the importance of unobservable characteristics based on the selection on observable characteristics. From the difference between the treatment effects in regressions with and without controls, it can be inferred how strong the selection on unobservables has to be to bring the treatment effect to zero. If the parameter that describes the relationship between selection on observables and unobservables, δ, is negative, this means that the selection on unobservables would need to have the opposite sign than the selection on observables, such that the unobservables would lead to a downward bias in the estimates. In Appendix C.1, we perform this analysis for all eleven outcomes, and constantly obtain negative selection parameters. This suggests that selection on unobservables, if at all, would have to go in the opposite direction to bring the observed treatment effect to zero. This lends support to our identifying assumption that the sample is balanced between people living in high- and low-exposure areas.

7.1.2 Instrumental variables estimation

To address concerns about the exogeneity of the ground deposition of Cs137 in May 1986 in our regression, we perform an instrumental variable (IV) estimation. We exploit that the level of fallout in an area depends on the level of rainfall within a critical time window of ten days in late April and early May 1986. To instrument for the ground deposition, we use deviation in rainfall from the usual rainfall levels in a municipality in late April and May 1986. In both the first and the second stage, we control for the average level of rainfall in the period from April 29 to May 8 in the years 1981 to 1985. This instrument provides us with a strong first stage ($F = 143.1$). At the same time, it plausibly fulfills the exclusion restriction that idiosyncratic rainfall on five days in 1986 have no direct impact on test scores 25 years later.

Besides addressing the potential endogeneity of the regressor of interest due to residential sorting or other forms of self-selection, the instrumental variable approach also addresses measurement error. Given that we only observe a person's potential but not actual exposure, our regressor of interest is likely measured with error. To the extent that the level of rainfall is uncorrelated with the measurement error in the fallout — which is plausible given that measuring points for rainfall are in different locations than those measuring radiation — the instrumental variable approach can eliminate the bias from measurement error.

Table 7 displays the instrumental variable estimates for all eleven outcomes as well as the OLS results for comparison. In all regressions, we control for individual and municipality characteristics as well as state fixed effects. All IV estimates confirm the negative sign found in the OLS estimates. The IV estimates are substantially larger than the OLS estimates. This difference in magnitude points to measurement error in the exposure to radiation, which may lead to a downward bias in the OLS estimates. An additional explanation for this difference could be the the difference between the local average treatment effect (LATE) estimated by the IV and the average treatment effect estimated by the OLS. The IV estimate is identified by compliers, that is, those parts of the German population whose cognitive skills are particularly responsive to radiation. Overall, the IV results, while less precisely estimated than the OLS counterparts, confirm the negative impact of radiation on cognitive skills, and suggest that our baseline estimates in Section 6 are on the conservative side.

7.1.3 Regressions with grid-level fixed effects

An alternative way to reduce the influence of unobserved heterogeneity is to compare people that live relatively close to each other, which presumably have more similar characteristics and are subject to the same institutions compared to people living further apart. In our case, a complication with the comparison of people living close together is that our regressor is measured at the municipality-level. The next highest administrative unit would be a county, although there is too little variation across municipalities within counties to allow for precise estimation of β.

As an alternative, we lay a grid of $120 \times 120km$ cells over the map of Germany, and estimate a model with grid cell fixed effects. This approach exploits variation within grid cells, thereby comparing people with the same characteristics living in the same grid cell but differentially exposed to radiation. To account for the arbitrary locus of the grid, we perform 500 replications of the same procedure, whereby in each replication we shift the grid by random distances north-south and east-west.[17]

Figure 3 plots the empirical distributions of the estimates for all eleven outcomes. The results are in line with our baseline results obtained in Section 6. For some outcomes such as math, reading and listening, the estimates are large and the entire distribution of estimates lies below zero. For other outcomes, such as reasoning or reading speed, the estimates are centered

[17] This approach follows Barsbai et al. (2017). We are grateful to Andreas Steinmayr for sharing his code.

Table 7: Instrumental variable estimates

	(1) OLS	(2) IV
A. Individual test scores		
Math	-0.011***	-0.038
	(0.003)	(0.025)
Reading	-0.014***	-0.046*
	(0.005)	(0.025)
Listening comprehension	-0.009**	-0.039
	(0.004)	(0.024)
ICT	-0.006	-0.006
	(0.003)	(0.019)
Scientific literacy	-0.004	-0.037*
	(0.003)	(0.020)
Reasoning	-0.001	-0.044*
	(0.004)	(0.024)
Reading speed	-0.008**	-0.009
	(0.004)	(0.022)
Perceptual speed	-0.004	-0.023
	(0.003)	(0.022)
B. Indices		
Cognitive skill index	-0.008***	-0.033*
	(0.003)	(0.018)
Crystallized intelligence index	-0.009***	-0.038**
	(0.003)	(0.018)
Fluid intelligence index	-0.006*	-0.023
	(0.003)	(0.020)
First-stage: dep. var. Cs137 kBq/m^2		
Precipitation (mm/m^3)	1.420***	
	(0.118)	
F statistic	143.555	

Notes: This table displays the results of separate regressions of the outcome variables listed on the left on the ground deposition of Cs137. All regressions control for state fixed effects, individual and municipality characteristics, which include a control for the average rainfall between April 29 and May 8 in the years 1981-1986. In Column (1), the ground deposition has been instrumented with the precipitation level on the same days in 1986. The results in Column (2) correspond to the baseline results in Table 3, Column (8). Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

around zero. In Appendix C.2, we additionally report the average p-value for each outcome, which provides a robustness check to our inference.

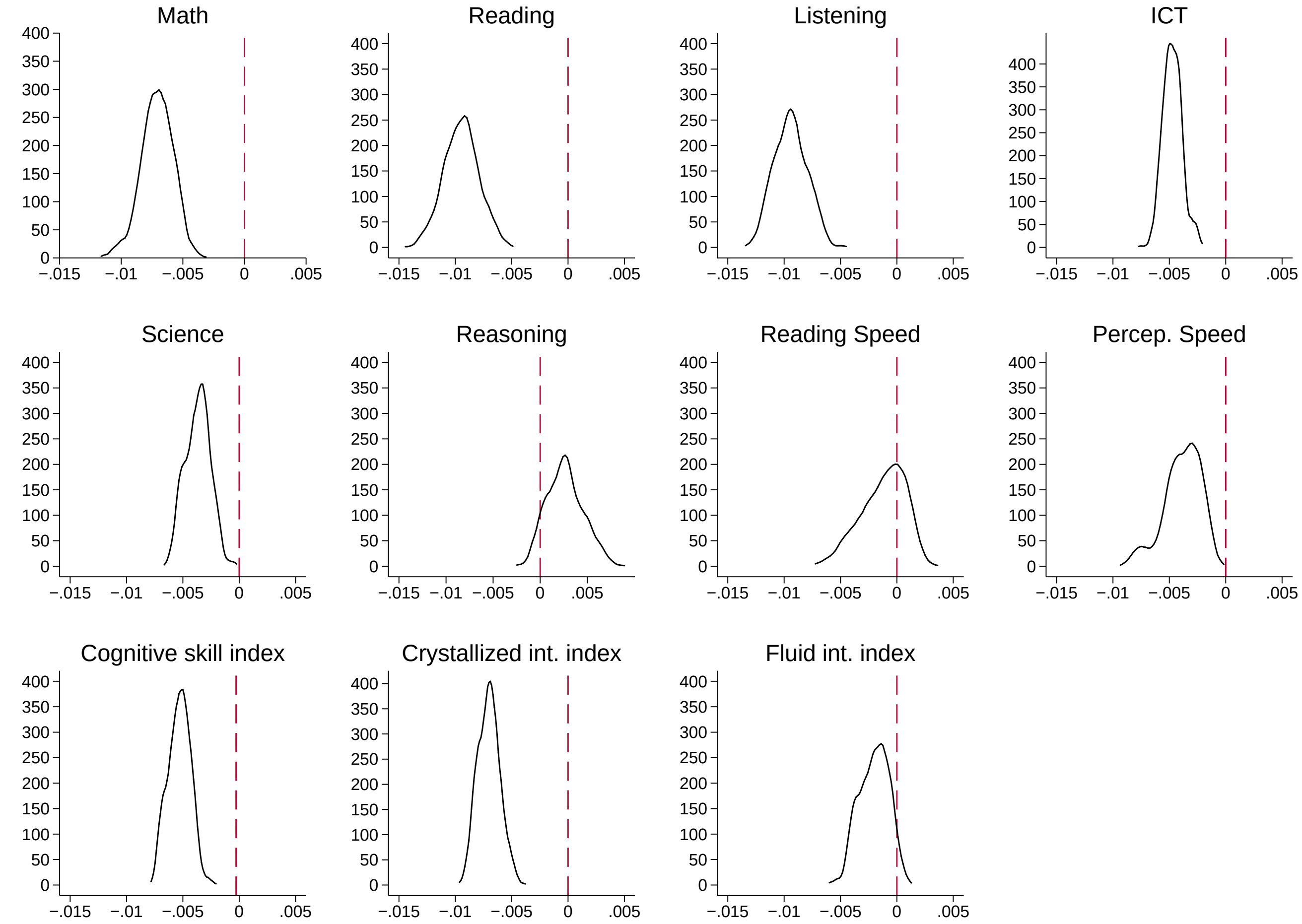


Figure 3: Empirical distribution of fixed effect estimates with random grid cells

Notes: This figure displays the empirical distributions of regression coefficients $\hat{\beta}$ with fixed effects at the level of $120 \times 120km$ grid cells. Each regression is estimated 500 times; in each replication, the grid is shifted by random amounts horizontally and vertically. The vertical lines indicate an effect of zero. In all regressions, we control for individual and municipality characteristics.

7.2 Addressing measurement error

Our estimates are potentially biased if radiation exposure is measured with error. This is almost certainly the case because our regressor measures the average potential exposure rather than the actual exposure. After Chernobyl, every person in Germany received an individual dose of radiation that depends on the person's exact location of residence in April 1986 as well as his or her lifestyle. Our regressor, in contrast, measures the initial potential exposure of the average person in a given municipality. We addressed this type of measurement error with the instrumental variable approach in Section 7.1.2, and the results point to attenuation bias.

A second source of measurement error is the linkage of radiation data via the centroid of a municipality. The geographic center of a municipality can be in a different location compared to the population center, such that the average person is in fact exposed to a different level of radiation compared to the one at the geographic center. To address this concern, we perform a series of eight robustness checks, whereby in each robustness check we link the data based on different points. For example, we link the data based on the population center or the point with the highest population density, or compute the population-weighted average radiation based on the radiation data at a $3 \times 3km$ grid cell level. The results, shown in the appendix, turn out to be robust to the linkage procedure, suggesting that the measurement error from the data linkage is small.

7.3 Selective attrition

Selective mortality There are several sources of attrition which, if systematic, can bias our estimates. One potential source of attrition is selective mortality. Radioactivity is known to increase the risk of developing cancer, which in general is one of the major death causes in Germany. If radiation induces a higher risk of dying from cancer among older workers, our sample may no longer be representative of the affected population, as it only comprises the survivors. To test if selective mortality is an issue, we obtained county-level mortality data from the life tables provided by the German statistical office. Based on a regressions of county-cohort-year-specific mortality rates on the county-level ground deposition of Cs137 and state fixed effects, we find no evidence for selective mortality. The results are described in greater detail in Appendix D.

Survey-design-based attrition In Appendix D, we further test for systematic attrition due to the sample design or missing information. Based on a regression with municipality-level controls and state fixed effects, we show that the ground deposition of Cs137 neither predicts non-participation in the competence test nor attrition due to missing personal information. Furthermore, we test whether the fact that not every respondent took every competence test is related to radiation exposure and find no evidence thereof. Finally, the sampling procedure of NEPS SC6 randomly sampled people from 250 German municipalities. Our results show that, in line with random sampling, the ground deposition of Cs137 does not predict the inclusion into the sample.

Constructing a standardized index in the presence of attrition. One of the outcomes reported with our baseline results is a standardized index of all eight cognitive test scores. However, as briefly explained in Section 4, not every person in the sample completed every test. While there appears to be no selective attrition from each test, attrition presents a challenge for the construction of the standardized index and the interpretation of the results. For the estimation in Section 6, we construct the standardized index based on all test scores a person obtained; if someone completed all eight tests, the index is the average score of these eight tests; if someone completed one test, the index is this one test score.

To test for the robustness of our results to the construction of the index in the presence of attrition, we estimate separate regressions with samples restricted to people who completed at least a number of tests. The results are reported in Appendix D.3. For example, out of 4,423 observations, 1,029 completed all eight tests, 3,179 completed at least five tests, and 4,417 at least seven tests. The estimates are larger the more tests people completed, ranging from -0.013 for those with all eight tests to -0.008 for those with at least one test. All results are statistically significant at the 5%-level or lower. This suggests that, if anything, by using all available test scores to construct the index, we under-estimate the true effect.

7.4 Inference

Multiple hypothesis testing In our main analysis, we present separate regression results for a large number of outcomes. The outcomes, measuring different dimensions of the latent factor cognitive skills, are correlated across observations. For example, if radiation has a negative effect on reading scores, it most likely also has a negative effect on math scores. If we run separate regressions for each outcomes, the conventional test statistics do not take this correlation into account, resulting in an over-rejection of the null hypothesis. Therefore, the more hypotheses are tested, the higher is the chance of obtaining at least one statistically significant coefficient.

The statistical literature proposes two approaches to overcome the multiple testing problem.[18] One is to keep the number of tests constant but adjust the p-values and t-statistics of each hypothesis test to account for the correlations between outcomes. The other is to leave the p-values and t-statistics unadjusted but reduce the number of hypotheses to one by using a standardized index as outcome.

In Appendix E.2, we report the results of both approaches. To adjust the p-values, we use the step-down procedure developed by Benjamini and Hochberg (1995). This procedure first sorts the p-values of all hypotheses tests from smallest to lowest, and adjusts each p-value depending on its rank position. The smallest p-value receives the largest adjustment while the largest p-value does not get adjusted at all.[19] After the correction of the p-values, the effects on three outcomes, namely math, reading and listening, remain statistically significant at the 5%-level, and the one on reading speed at the 10%-level.

As an alternative, we perform a summary index test whereby all outcomes are summarized in a standardized index. Following OB́rien (1984) and Anderson (2008), we construct the index by standardizing each outcome to mean zero and standard deviation one and building for each

[18] See Anderson (2008) for a discussion and applications.

[19] An alternative that is often used is the Bonferroni correction, which simply multiplies each p-value by the number of hypothesis tests. However, as shown by Benjamini and Hochberg (1995), the Bonferroni correction is very conservative and leads to a severe under-rejection of the null hypothesis, while the step-down procedure provides a more accurate adjustment of p-values.

observation a weighted sum of the standardized outcomes. The weights are constructed such that outcomes that are highly correlated — and, therefore, add little new information to the index — receive lower weight than outcomes with a low correlation that add more new information.[20] Using the weighted index, the estimated effects are similar to those on the standardized index reported in Section 6, both in terms of point estimates and standard errors.

Randomization inference A second issue with inference is that hypothesis tests reported in Section 6 are based on conventional p-values and t-tests and their underlying parametric assumptions. These assumptions may not be valid, for example because the error terms are correlated across individuals. So far, we adjust for cross-sectional dependence of error term by clustering the standard errors at the county-level. However, it is unclear if clustering at this level leads to a sufficient adjustment of the standard errors.

To gain further confidence in the statistical significance of our results, we perform a series of non-parametric permutation tests. The idea behind these tests is to obtain the sampling distribution of estimates under the assumption that the null hypothesis of no effect is true. This placebo distribution can be obtained in two ways, namely by repeatedly randomizing the treatment across observations while leaving the outcome and all other regressors constant, or by repeatedly randomizing the outcome while leaving the treatment and all other regressors constant. The null hypothesis can be rejected if the original point estimate lies in the tails of the placebo distribution and, thus, is unlikely to emerge by chance. The tests, shown in Appendix E.1, provide strong evidence against the null hypothesis of no effect. When we randomize the ground deposition across all observations and run 10,000 replications, we obtain an empirical p-value of 0.0004, whereas if we randomize within federal states, we obtain a p-value of 0.0032.

8 Conclusion

In this paper, we use geographic variation in radioactive ground contamination to explore how exposure to subclinical levels of radiation negatively affects people's cognitive skills. Focusing on Germany, we show that people who lived in areas that received high levels of radioactive rainfall immediately after the Chernobyl nuclear disaster in 1986 performed significantly worse in cognitive tests more than 25 years later. The historical setting — Germans only learned about the disaster after the rainfall occurred — allows us to circumvent a major threat to identification, namely ex-ante avoidance behavior. Moreover, we find no evidence of systematic residential sorting. Combined with extensive placebo and permutation tests, these facts suggest that our results have a causal interpretation.

Given the importance of cognitive skills for many life outcomes, our findings suggest that

[20] Technically, the weights are constructed from the row sums of the covariance matrix of all outcomes, whereby the elements on the diagonal are set to zero. An outcome that has a higher correlation with all others has a higher row sum. This inverse of this sum is used as weight. See Appendix E.2 for further details.

subclinical radiation imposes non-trivial social costs. While it is impossible for anyone to escape radiation altogether, public policies can help to reduce exposure, for example through regulations in medical treatment, building regulations, workplace regulations and by substituting nuclear power for other sources of energy generation. Policymakers will need to weigh the costs of such regulations against the human capital cost of higher radiation. With returns to human capital having considerably increased in recent years, the social costs will increase as well.

An important task for future research is to disentangle the biological effects of radiation exposure from compensating behavior. Our reduced-form estimate includes both channels. Given that we find evidence for compensating behavior, it is most likely a lower bound to the biological effect. Identification of the pure biological effect, however, would require an empirical setting in which people cannot engage in compensating behavior, as well as data on brain activity. In addition, it would be important to assess which behaviors — changes in nutrition, modifications to buildings, etc — are most effective in counteracting the negative effect of radiation.

References

Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives* 25: 153–172.

Almond, D., Edlund, L. and Palme, M. (2009). Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly journal of economics* 124: 1729–1772.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association* 103: 1481–1495.

Antoni, M., Drasch, K., Kleinert, C., Matthes, B., Ruland, M. and Trahms, A. (2011). Working and learning in a changing world: Part I: Overview of the study-March 2011 (Second, updated version). Tech. rep., Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].

Aust, F., Gilberg, R., Hess, D., Kleudgen, M. and Steinwede, A. (2011). Neps etappe 8 befragung von erwachsenen haupterhebung 1. welle 2009/2010 .

Barsbai, T., Rapoport, H., Steinmayr, A. and Trebesch, C. (2017). The effect of labor migration on the diffusion of democracy: Evidence from a former soviet republic. *American Economic Journal: Applied Economics* forthcoming.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57: 289–300.

Berendes, K., Weinert, S., Zimmermann, S. and Artelt, C. (2013). Assessing language indicators across the lifespan within the german national educational panel study (neps)/die erfassung von sprachindikatoren über die lebensspanne im nationalen bildungspanel. *Journal for educational research online* 5: 15.

Black, S. E., Bütikofer, A., Devereux, P. J. and Salvanes, K. G. (2013). This is only a test? long-run impacts of prenatal exposure to radioactive fallout. *National Bureau of Economic Research* .

Blossfeld, H.-P., Rossbach, H.-G. and Maurice, J. von (2011). Education as a lifelong process - the german national educational panel study (neps). *Zeitschrift für Erziehungswissenschaft* 14 [Special Issue].

Brenner, D. J., Doll, R., Dudley, T. G., Hall, E. J., Land, C. E., Little, J. B., Lubin, J. H., Preston, D. L., Preston, R. J. and Puskin, J. S. (2003). Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. *Proceedings of the National Academy of Sciences* 100: 13761–13766.

Bromet, E. J. and Havenaar, J. M. (2007). Psychological and perceived health effects of the chernobyl disaster: a 20-year review. *Health physics* 93: 516–521.

Brunner, M., Lang, F. R. and Lüdtke, O. (2014). Erfassung der fluiden kognitiven Leistungsfähigkeit über die Lebensspanne im Rahmen der National Educational Panel Study: Expertise [Assessment of fluid cognitive skills over the life span in the National Educational Panel Study: Expertise]. Tech. rep., NEPS Working Paper 42.

Bundesamt für Strahlenschutz (2016). Der reaktorunfall 1986 in tschernobyl .

Bundesregierung (1986-1991). Bericht der bundesregierung über umweltradioaktivität und strahlenbelastung. *Deutscher Bundestag* .

Bunzl, K., Schimmack, W., Krouglov, S. and Alexakhin, R. (1995). Changes with time in the migration of radiocesium in the soil, as observed near chernobyl and in germany, 1986–1994. *Science of the total environment* 175: 49–56.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*, *35*. Elsevier.

Clark, M. and Smith, F. (1988). Wet and dry deposition of chernobyl releases. *Nature* 332: 245–249.

Danzer, A. M. and Danzer, N. (2016). The long-run consequences of chernobyl: Evidence on subjective well-being, mental health and welfare. *Journal of Public Economics* 135: 47–60.

Deng, W., Aimone, J. B. and Gage, F. H. (2010). New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nature reviews. Neuroscience* 11: 339.

Douw, L., Klein, M., Fagel, S. S., Heuvel, J. van den, Taphoorn, M. J., Aaronson, N. K., Postma, T. J., Vandertop, W. P., Mooij, J. J. and Boerman, R. H. (2009). Cognitive and radiological effects of radiotherapy in patients with low-grade glioma: long-term follow-up. *The Lancet Neurology* 8: 810–818.

Drasch, K. and Matthes, B. (2013). Improving retrospective life course data by combining modularized self-reports and event history calendars: experiences from a large scale survey. *Quality & Quantity* : 1–22.

Fielitz, U. and Richter, K. (2013). Bundesweiter überblick über die radiocäsiumkontamination von wildschweinen-vorhaben 3607s04561 .

Gehrer, K., Zimmermann, S., Artelt, C. and Weinert, S. (2012). The assessment of reading competence (including sample items for grade 5 and 9. *Leibniz Institute for Educational Trajectories* .

Gehrer, K., Zimmermann, S., Artelt, C. and Weinert, S. (2013). Neps framework for assessing reading competence and results from an adult pilot study/neps-rahmenkonzeption zur messung von lesekompetenz und resultate einer pilotstudie mit erwachsenen. *Journal for Educational Research Online* 5: 50.

Gonzalez, A. B. de and Darby, S. (2004). Risk of cancer from diagnostic x-rays: estimates for the uk and 14 other countries. *The lancet* 363: 345–351.

Haberkorn, K. and Pohl, S. (2013). Cognitive basic skills (non verbal) data in the scientific use file. *University of Bamberg, National Educational Panel Study (NEPS),* .

Hachenberger, C., Trugenberger-Schnabel, A., Löbke-Reinl, A. and Peter, J. (2017). Umweltradioaktivität und strahlenbelastung–jahresbericht 2015. *Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU)* .

Hahn, I., Schöpss, K., Saß, S., Hansen, S., Martensen, M., Wagner, H. and Funke, L. (2014). The assessment of scientific literacy in the national educational panel study (neps), including example items for kindergarten, grade 6, students and adults. *Leibniz Institute for Educational Trajectoriestute for Educational Trajectories* .

Hall, P., Adami, H.-O., Trichopoulos, D., Pedersen, N. L., Lagiou, P., Ekbom, A., Ingvar, M., Lundell, M. and Granath, F. (2004). Effect of low doses of ionising radiation in infancy on cognitive function in adulthood: Swedish population based cohort study. *British Medical Journal* 328: 19.

Ham, A.-K. van den, Ehmke, T., Hahnand, I., Wagner, H. and Schöps, K. (2016). Mathematische und naturwissenschaftliche Kompetenz in PISA, im IQB-Ländervergleich und in der National Educational Panel Study (NEPS) Â–Vergleich der Rahmenkonzepte und der dimensionalen Struktur der Testinstrumente. In *Forschungsvorhaben in Ankopplung an Large-Scale-Assessments*, 140–160.

Hendee, W. R. and O'Connor, M. K. (2012). Radiation risks of medical imaging: separating fact from fantasy. *Radiology* 264: 312–321.

Hou, X., Fogh, C. L., Kucera, J., Grann, K. A., Dahlgaard, H. and Nielsen, S. P. (2003). Iodine-129 and caesium-137 in chernobyl contaminated soil and their chemical fractionation. *Science of the total environment* .

Ihme, J. M., Senkbeil, M. and Wittwer, J. (2015). The neps ict literacy framework and item examples. *Leibniz Institute for Educational Trajectoriestute* .

International Atomic Energy Agency (2006). Environmental consequences of the chernobyl accident and their remediation: twenty years of experience. report of the chernobyl forum expert group environment. *International Atomic Energy Agency* .

Lehmann, H. and Wadsworth, J. (2011). The impact of chernobyl on health and labour market performance. *Journal of health economics* 30: 843–857.

Leuraud, K., Richardson, D. B., Cardis, E., Daniels, R. D., Gillies, M., O'Hagan, J. A., Hamra, G. B., Haylock, R., Laurier, D., Moissonnier, M., Schubauer-Berigan, M. K., Thierry-Chef, I. and Thierry-Chef, I. (2015). Ionising radiation and risk of death from leukaemia and lymphoma in radiation-monitored workers (inworks): an international cohort study. *The Lancet Haematology* 2(7): 276–281.

Little, M. P., Wakeford, R., Tawn, E. J., Bouffler, S. D. and Berrington de Gonzalez, A. (2009). Risks associated with low doses and low dose rates of ionizing radiation: why linearity may be (almost) the best we can do. *Radiology* 251: 16–12.

Mettler, F. A., Huda, W., Yoshizumi, T. T. and Mahesh, M. (2008). Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology* 248: 254–263.

Møller, A. and Mousseau, T. A. (2013). The effects of natural variation in background radioactivity on humans, animals and other organisms. *Biological Reviews* 88: 226–254.

Monje, M. and Dietrich, J. (2012). Cognitive side effects of cancer therapy demonstrate a functional role for adult neurogenesis. *Behavioural brain research* 227(2): 376–379.

Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E. and Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan/modellierung und erfassung mathematischer kompetenz über die lebensspanne. *Journal for Educational Research Online* 5: 80.

OB́rien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* 4: 1079–1087.

OECD (2016). Radiological protection science and application. *OECD* .

Ortwin, R. (1990). Public responses to the chernobyl accident. *Journal of Environmental Psychology* 10: 151–167.

Oster, E. (2014). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* .

Reimer, M. and Matthes, B. (2007). Collecting event histories with truetales: Techniques to improve autobiographical recall problems in standardized interviews. *Quality & Quantity* 41: 711–735.

Rola, R., Raber, J., A., R., Otsuka, S., VandenBerg, S. R., Morhardt, D. and Fike, J. R. (2004). Radiation-induced impairment of hippocampal neurogenesis is associated with cognitive deficits in young mice. *Experimental neurology* 188(2): 316–330.

Salthouse, T. (2012). Consequences of age-related cognitive declines. *Annual review of psychology* 63.

Schnittjer, I. and Duchhardt, C. (2015). Mathemathcal competence:framework and exemplary test items. *Leibniz Institute for Educational Trajectories* .

Senkbeil, M. and Ihme, J. M. (2015). NEPS Technical Report for Computer Literacy: Scaling Results of Starting Cohort 6–Adults. Tech. rep., NEPS Working Paper.

Senkbeil, M., Ihme, J. M. and Wittwer, J. (2013). The test of technological and information literacy (tilt) in the national educational panel study: Development, empirical testing, and evidence for validity/test zur erfassung technologischer und informationsbezogener literacy (tilt) im nationalen bildungspanel: Entwicklung, empirische überprüfung und validitätshinweise. *Journal for educational research online* 5: 139.

Spalding, K. L., Bergmann, O., Alkass, K., Bernard, S., Huttner, M. S. A. B., Boström, E., Westerlund, I., Vial, C. and Buchholz, B. A. (2013). Dynamics of hippocampal neurogenesis in adult humans. *Cell* 153(6): 1219–1227.

Squire, L. R. (2009). The legacy of patient hm for neuroscience. *Neuron* 61(1): 6–9.

Stewart, F., Akleyev, A., Hauer-Jensen, M., Hendry, J., Kleiman, N., Macvittie, T., Aleman, B., Edgar, A., Mabuchi, K. and Muirhead, C. (2012). Icrp publication 118: Icrp statement on tissue reactions and early and late effects of radiation in normal tissues and organs–threshold doses for tissue reactions in a radiation protection context. *Annals of the ICRP* 41: 1–322.

Supekar, K., Swigart, A. G., Tenison, C., Jolles, D. D., Rosenberg-Lee, M., Fuchs, L. and Menon, V. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. *Proceedings of the National Academy of Sciences* 110(20): 20–27.

Tubiana, M., Feinendegen, L. E., Yang, C. and Kaminski, J. M. (2009). The linear no-threshold relationship is inconsistent with radiation biologic and experimental data. *Radiology* 251: 13–22.

UNSCEAR (1994). Sources, effects and risks of ionizing radiation: Unscear 1994 report. *United Nations Scientific Committee on the Effects of Atomic Radiation and others* .

UNSCEAR (2000). Sources and effects of ionizing radiation. *United Nations Scientific Committee on the Effects of Atomic Radiation and others* Volume I: SOURCES.

UNSCEAR (2008). Sources and effects of ionizing radiation. *United Nations Scientific Committee on the Effects of Atomic Radiation and others* .

UNSCEAR (2014). Sources, effects and risks of ionizing radiation: Unscear 2013 report. *United Nations Scientific Committee on the Effects of Atomic Radiation and others* .

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T. and Carstensen, C. H. (2011). Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft* 14: 67–86.

Wetterdienst, D. (2017). Richtlinie: Automatische nebenamtliche wetterstationen im dwd .

Winkelmann, I., Endrulat, H., Fouasnon, S., Gesewsky, P., Haubelt, R., Klopfer, P., Köhler, H., Kohl, R., Kucheida, D., Müller, M. et al. (1986). Ergebnisse von radioaktivitätsmessungen nach dem reaktorunfall in tschernobyl. *ISH-99, Institut für Strahlenhygiene des Bundesgesundheitsamtes, Neuherberg* .

Winkelmann, I., Haubelt, R., Neumann, P. and Fields, D. (1989). Radionuclide deposition and exposure in the Federal Republic of Germany after the Chernobyl accident. Tech. rep., Oak Ridge National Lab., TN (USA).

Yasunari, T. J., Stohl, A., Hayano, R. S., Burkhart, J. F., Eckhardt, S. and Yasunari, T. (2011). Cesium-137 deposition and contamination of japanese soils due to the fukushima nuclear accident. *Proceedings of the National Academy of Sciences* 108: 19530–19534.

Zimmermann, S., Artelt, C. and Weinert, S. (2014). The assessment of reading speed in adults and first-year students .

Online Appendices

(Not for publication)

A Data description

A.1 Sampling in the ALWA subsample

A Data description A.1 Sampling in the ALWA subsample As described in Section 3, our main data source is the ALWA subsample of the NEPS Adult Cohort (SC6). Here we provide more detailed information on the sampling procedure. ALWA was sampled in two steps. First, weights of municipalities were assigned based on the residential population of the cohorts born between 1956 to 1986. There was no minimum number of residents per municipality, such that all municipalities were included. From all German municipalities, a random sample of 250 was drawn, from which potential participants were included in the survey. Second, individuals were randomly selected from person registers within municipalities. The number of selected individuals per municipality was set according to the municipality's weight, resulting in a sample of 42,712 addresses that are representative of these age cohorts in Germany. For all selected individuals, telephone numbers were collected and participants were contacted by phone. The telephone number for 22,656 of these was identified. Out of these, 10,404 interviews could be realized between August 2007 and April 2008, leading to response rate of 24.4/Before receiving the first call attempt, participants were sent information material about the study. Furthermore, to increase the willingness to participate, material incentives were provided; among all participants, 60 prizes, such as laptops, travel vouchers, or iPods were distributed through a lottery (Antoni et al., 2011). Computer-assisted telephone interviews (CATI) were used to collect information about current personal characteristics and about past events regarding residential, occupational and educational history.

To collect the residential information, interviewers asked participants to state the name of the municipality of residence. If a person lived abroad, the name of the country of residence was collected. Municipality lists were provided to interviewers in order to enable a precise assignment of municipalities. In cases where municipality names were identical, interviewers asked about the county or federal state. Municipality keys were assigned by the interviewer based on the definition of 2004, although for the current NEPS datasets the municipality keys have been transformed to the definition of 2013.

To minimize recall problems, the interviewers used a survey technique called TrueTales, which enhances respondents' memory based on the interconnection of modularized self-reports and event history calendars (EHC) (Reimer and Matthes, 2007). Key to this technique is that participants go through each domain of their life history — education, residence, and work — separately. The interview process does not follow a continuous time line, but is rather based on events in a person's history, such as going to school, finishing college, or getting married. This procedure enhances participants' autobiographical memory. In addition, interviewers used a computer software that pointed out spatial as well as chronological inconsistencies between the three domains (Drasch and Matthes, 2013).

Each module starts with a respondent's birth and further goes through their lives. In case of the residential history, participants stated the current name of the municipality the residence was located in. Participants could state the municipality of their primary and secondary residency, although we only focus on the primary residence. In the education module, participants were asked to state the place and the type of educational institution they attended during a given spell. The employment module contains information about the employer, such as the location or sector, as well as contractual details such as the type of employment, income and working hours.

A.2 Competence tests

A.2.1 Further details on test scores

NEPS is designed to assses competence development accross the lifespan starting with Newborns SC1, over pupils SC2-SC4 and Students SC5 to adults SC6. All cohorts are testet in the same dimensions and tests are strongly orientated towards the concepts used by PISA. However, in order to compare results of children with those of adults, some adjustments are necessary leading to deviations from concepts used by PISA. Furthermore, the necassity of comparable test for children and adults explain the greater proximity to PISA then to PIAAC, for example. We will explain the construction of all test dimensions covered in the SC6 in the following.

Reading competence The Assessment of *reading competence* in the SC6 of NEPS includes different text functions like literary texts or advertising texts where participants are required to identify information, draw test-related conclusions and find the core message of the text. The maximum test score equals 39 points. The maximum processing duration is 28 minutes by paper-pencil questionnaires (Gehrer et al., 2012).

Functional understanding is the basis for the concept of reading competence in the SC6 of NEPS. It focuses on competent handling of written texts in typical everyday situations. This orientation draws on the concept of literacy in international studies of reading competence — such as the International Adult Literacy Survey (IALS), or the multicycle comparisons of school performance in the Programme for International Student Assessment (PISA) — with a focus on enabling participation in society.

However, the concept of reading competence in NEPS is distinguished from PISA due to two main reasons. In international studies of reading competence (e.g. PISA, CEFR), underlying texts are often cathegozized according to the type of situation in which they are applied - commonly with a focus on the reasons for reading like education, work, the personal domain and the public domain. However, reading competence in NEPS is less orientated toward the reasons for reading, but focuses predominantly on the functions of text along with the types of text associated with these functions and how these relate to the cognitive requirements of reading. Furthermore, whereas PISA ueses discontinuous texts NEPS does not. Continuous texts exclusively transport verbal information in the form of letters. Discontinuous texts extend this by linking the written verbal information to pictorial information like tables, graphs or diagrams. The combination of continuous and discontinuous texts results in a broader concept of reading competence. As a result, the concept of reading competence in NEPS requieres slightly different cognitive requirements than the concept used in PISA, shown by tests measuring external validity (Gehrer et al., 2013).

Mathematical competence The test of Mathematical competence consist in the SC6 in NEPS consists of 21 items. Each is item equivalent to one point in the test score. The maximum processing duration is 28 minutes by paper-pencil questionnaires (Schnittjer and Duchhardt, 2015).

In order to be compatible with literacy view of mathematical competence in PISA, the test of mathematical competence in the SC6 of NEPS has been developed in very close connection to the the PISA framework. Thus, measures reveal the ability to flexibly use and apply mathematics in realistic daily situations requiring mathematical skills like systematic trying or generalizing and mathematical knowledge like known algorithms or calculation methods. Therefore, it does not describe the outcomes of mathematics teaching — imply consequences for the curriculum—, but rather required abilities and skills of daily lives.

Like in PISA mathematical cometence in the SC6 comprisis four content areas which require six cognitive processes. Content areas are quantity, change and relationships, space and shape, and data and chance. The six included cognitive processes are mathematical communication, mathematical argumentation, modeling, using representational forms, mathematical problem solving, and technical abilities and skills (Neumann et al., 2013). First test on external validity indicate comparability with dimensions measured by PISA (van den Ham et al., 2016).

Scientific literacy The concept of *scientific literacy* follows the concept of the American Association of Advancement of Science (AAAS) and PISA. Using 22 items this tests describes individual knowledge of basic scientific concepts and facts (KOS) —divided into the content-related components matter, systems, development and interactions— and the understanding of scientific processes (KAS) — divided into the process-related components scientiffic enquiry and scientiffic explanations — required for personal decision making. The maximum attainable test score is 28 points. The maximum processing duration is 25 minutes by paper-pencil questionnaires.

As in the PISA framework areas (KAS) and (KOS) are implemented in the contexts area health, environment and technology. The concept of scientific literacy in NEPS is slightly distinguished from PISA because contexts do not explicitly consider the PISA perspective of personal, social and global situations. Situations rather occur implicitly and thus cannot be analyzed separately because each situation is represented by a limited number of items. Time constrains limit the number of items. Test itmes are organized in units combining context and component (KAS or KOS). This approach is similar to PISA. However, different than in PISA not all context-compoment combinations could be tests due to time constrains. Nevertheless, every component is measured in at least two contexts. First test on external validity indicate comparability with dimensions measured by PISA (Hahn et al., 2014).

Listening comprehension This test analysis receptive vocabulary. It measures the individual spectrum of vocabulary using spoken language by speakers. Participants face 89 items where they have to assign heard words to a sample of four pictures in front of them. Hence, the maximum attainable test score is 89 points by paper-pencil questionnaires . It follows the concept of the Peadbody Picture Vocabulary Test (PPVT) which is for example used in the British Cohort Study or the European Child Care and Educaton Study (ECCE). For the SC6, NEPS uses the public available German version of the PPVT published in 2004 (Berendes et al., 2013).

Information and Communication Technologie Literacy *Information and Communication Technologie Literacy* includes components of computer literacy representing knowledge and skills which are necessary for problem-oriented use of modern information technology. This entails knowledge about basic operations, create and edit documents as well as find and assess information. This is in line with the literacy concept of PISA. The maximum test score is 68 points within 25 minutes by paper-pencil questionnaires (Ihme et al., 2015). The tests exists of 29 items. There are two types of response formats: A complexe multiple choice format and a simple multiple choice format. Most items use screenshoots of spreadsheets, for example. ICT literacy in NEPS uses the framework "Digital transformation. A frameowrk for ICT literacy" of the International ICT Literacy Panel. It identifies four process components: access, creat, manage and evaluate. Acess and creat are cathegorized into technological literacy. Manage and evalute are cathergorized into information literacy. Software applications used to locate, process, present and communicate information can be categorized into word processing, spreadsheets, presentation and graphic software, e-mail and other communication applications (e.g. forums) and internet-based search engines and databases. Itmes are equally distributed across the four process components and four software applications. However, operationalizing of ICT via paper-pencil is not optimally because it can not capture complexe processes like organizing or structuring information (Senkbeil et al., 2013) (Senkbeil and Ihme, 2015)

Reading speed The assessment of *reading speed* in the NEPS, capturing basic reading processes as decoding, lexical access and basic sentence processing, consists of 51 short and simple sentences which have to be rated true or false (e.g. Mice can fly). Thus, the tests focusses on automatized reading processes. The maximum attainable test score is 51. Given the aim of assessing automized reading processes the sentence draw only in common world knowledge. Participants have two minutes time to rate as many sentences as possible by paper-pencil questionnaires. The test is based on the principles of the Salzburg reading screening (SLS) (Zimmermann et al., 2014). Information on external validity does not exisit.

Perceptual speed The test on *perceptual speed* reveals cognitive basic skills or more precisely the basal speed of information processing using picture symbol tests. The picture symbol test exists of two tables where in one table each graphical symbol has a specific number. In the second table the exact symbols exists as well, however, the corresponding number is missing. Participants have to find those numbers that equals the combination in the first table as fast as possible within 90 seconds with a maximum of 93 items by paper-pencil questionnaires. Each item equals one point in the test score. This procedere follows the digit symbol coding of the Wechsler Adult Intelligence test (Brunner et al., 2014) Information on external validity does not exisit.

Reasoning Another test for cognitive basic skills is a matrices test which covers *reasoning.* It consists of nine items with several horizontally and vertically arranged boxes in which different geometrical symbols are shown by paper-pencil questionnaires. One field is free which has to be filled base on a logical series. The maximum attainable test score is 12 points. This procedere follows the matrix reasoning of the Wechsler Adult Intelligence test (Brunner et al., 2014) (Haberkorn and Pohl, 2013). Information on external validity does not exisit.

Indexes Regarding the test dimensions available in NEPS within the framework of the literacy concept and based on the insights of the literature in section 2 we consider the psychological concept of fluid and crystallized intelligence (Cattell, 1987) as preferable to describe patterns of cognitive changes due to radioactive exposure. According to neurophysiological insights fluid and crystallized intelligence are functionally separate systems that have different neurological substrates.

This concept allows us to frame eight cognitive tests into two dimensions. Crystallized intelligence describes the cumulative products carried out in the past, in the form of acquired knowledge. Reasoning and novel problem solving as well as processing speed are considered to provide measures of fluid ability (Salthouse, 2012).

As different types of tests are expressed in different units, we convert the original scores to z-score units to facilitate comparisons across variables. We generate three indexes by accumulating existing tests. The first index *Cognitive skill index* is a sum of z-scores of every test performed by an individual between 2010 and 2015. Additionally, a *fluid intelligence index* is build covering processing speed as well as working memory based on the z-scores of reading speed, perceptual speed and reasoning. The remaining tests form the basis for an *crystallized intelligence index*, which can entail parts of fluid intelligence. Included z-scores of tests are mathematical competence, reading competence, information and communication technology literacy, scientific literacy and listening comprehension. In section 7 we will provide tests about the reliability of these indexes.

A.2.2 Participation in the competence tests

Tests in reading speed, math and reading were performed between October 2010 and Mai 2011. Between October 2012 and April 2013 test in ICT and scientific literacy were performed. Test in perceptual speed, listening comprehension and reasoning were performed between August 2014 and March 2015 . Most participants fulfilled their first in the first test period. However, some started in the second test period and a few only in the last period (Figure 4c). Figure 4d additionally shows that the majority performed at least seven tests. A negligible amount performed only one test. As shown in figure 4b people were assigned to four different Test groups which defined the test order. While the test order in the last wave between 2014 and 2015 was the same for every group, it differed in the first two test waves in 2010/2011 and 2012/2013. The test order in the first group was math, reading scientific literacy and it and the test order in the second group was reading, math, it and scientific literacy. Reading was skipped in the third and math in the fourth test group whereas the order stayed unchanged compared to group one or two where the corresponding test was performed first. Test groups were build to decrease panel attrition by lowering workloads for participants, whereas different test sequences ensure that the order does not play a role.

Figure 4a shows that participants do not necessarily perform the same tests. Most performed the reading speed test (3,602) whereas least performed the math test (2,644). As a result, 4,423 participants performed at least one test. Tests are randomly predetermined in different sequences and varying number of domains between 2010 and 2015 (Aust et al., 2011).

Participation in competence test was randomly selected. Survey participants either do not participate in competence tests or did a varying number of tests. The amount of tests was randomly selected as well varying between one and eight tests as shown in figure 4d.

According to Aust et al. (2011), some refused to participate in competence tests. This was especially true for less educated participants. Furthermore, older people slightly refuse participation more often. The test order and the amount of test slightly influence the propbability of realizing an interview. Performing two tests in the first wave intead of one reduces the probability a realizing the interview from 96.3% to 95.4%. Performing math first and than reading yield to 93.9% realized tests whereas performing reading and than math increased realization rate to 95.4% .

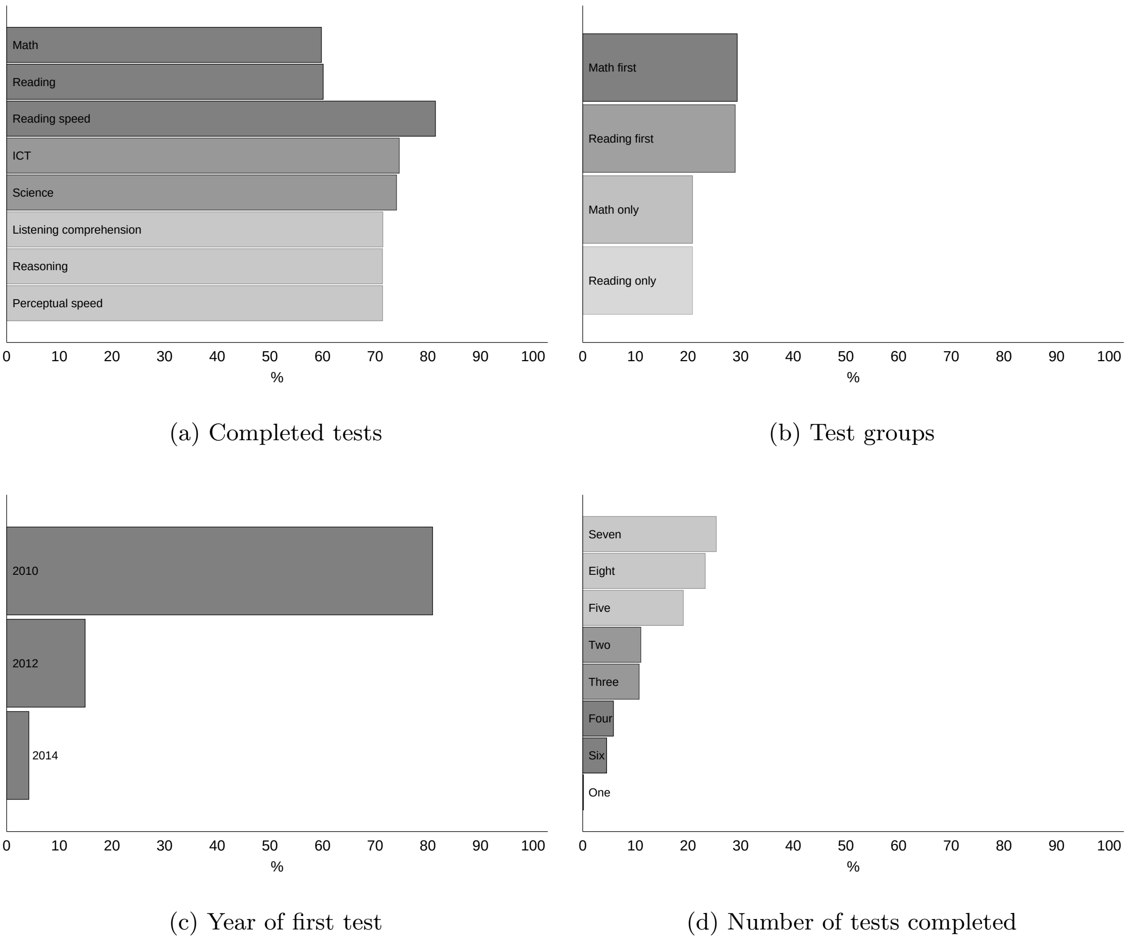
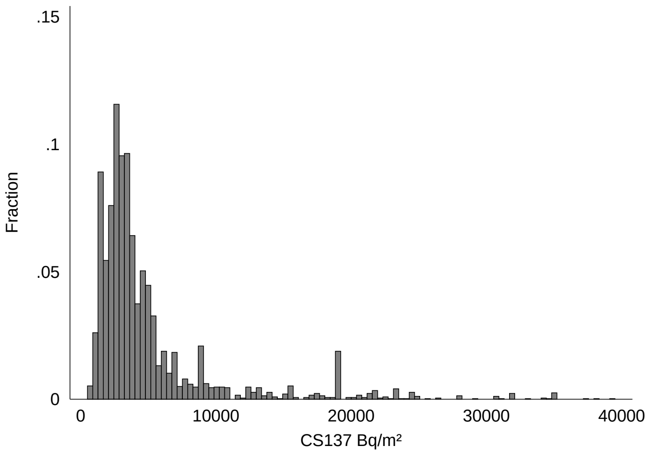


Figure 4: Participation in cognitive tests

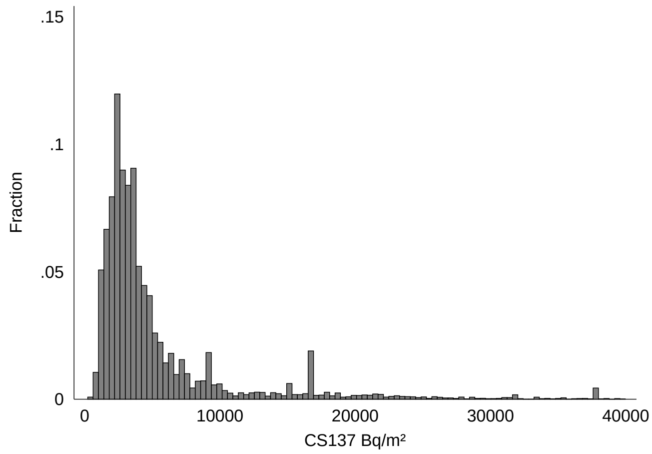
Notes: This figure displays descriptive statistics about the participation in cognitive tests for all 4,423 participants in our sample. Due to the survey design, not all participants took all tests, and tests were taken in different sequences. The first panel in the top row reports the share of participants who took each test. The second panel in the top row reports the distribution of test groups. The panels in the bottom row show the distribution across years in which the first test was taken, as well as the number of tests taken by each participant.

B Further Descriptive Statistics

Figure 5 displays the distribution of the fallout in our sample as well as the German population. Based on the municipality of residence in May 1986, Panel (a) lists for each person in our sample the ground deposition of Cs137 in Bq/m^2. Panel (b) shows the distribution for the entire German population, which we obtain by multiplying the ground deposition in each municipality with the population in 1997. The year 1997 was the first year for which consistent population data are available for the municipalities based on the same definition as the one used by NEPS.



(a) Sample



(b) Population

Figure 5: Variation in the ground deposition of Cs137in May 1986

Notes: This graph displays the distribution of the potential exposure to radiation, measured by the ground deposition of Cs137 in a person's municipality of residence in May 1986. Panel (a) displays the distribution in our sample, whereas Panel (b) displays the distribution in the German population. To obtain the distribution in the population, we computed the average ground contamination by municipality in 1986 and weighted the distribution by the population of each municpality in 1997. Sources: Federal Office for Radiation Protection (Bundesamt für Strahlenschutz) and The Service Center of the Federal Government for Geo-Information and Geodesy.

C Addressing unobserved heterogeneity

C.1 Selection on observables

One of the potential weaknesses of our analysis is omitted residential sorting. In order to cope with the potential effect of unobserved variables, we implement the method of (Oster, 2014). In Table 8 we analyze how much the coefficient of CS137 changes when we add controls and how important unobservables have to be for results to become insignificant, called delta. Higher values of delta imply more robust estimates. A negative delta means that if the observables are positively correlated with the treatment, the unobservables have to be negatively correlated with the treatment.

$$\delta \frac{\sigma_{1x}}{\sigma_1^2} = \frac{\sigma_2^x}{\sigma_2^2} \tag{4}$$

Values needed to produce an insignificant estimate of CS137 kBq/m^2 range between -0.029 and -1.375 if we restrict controls to federal state fixed effects in column (2). However, -0.029 refers to reasoning which is insignificant in our main analysis. For test that are highly significant in our main analyses like math and reading values are -0.515 or -1.045. Values for the cognitive skill index and the crystallized intelligence index are close to one, indicating that unobaservabels need to be explain almost as much as observable and need to be negativly correlated with the treatment, to get insignificant estimates of CS137 kBq/m^2. Thus, confounding effect of unobservables appear unlikely.

Table 8: Selection on unobservables

	(1)	(2)
A. Individual test scores		
Math	-0.190	-0.515
Reading	-0.342	-1.045
Listening comprehension	-0.465	-0.402
ICT	-0.282	-1.037
Scientific literacy	-0.192	-0.825
Reasoning	-0.192	-0.029
Reading speed	-0.067	-1.375
Perceptual speed	-0.168	-0.432
B. Indices		
Cognitive skill index	-0.274	-0.717
Crystallized intelligence index	-0.265	-0.846
Fluid intelligence index	-0.192	-0.370
Controls:		
State FE	No	Yes

Notes: This table displays the estimates for δ, the proportionality factor between selection on observable and unobservable characteristics. The calculation is based on the method proposed by Oster (2014) that compares the treatment effect of a restricted regression with and an unrestricted regression without controls. In Column (1), we condition in both regressions on state fixed effects, whereas in Column (2), the unrestricted regression is a bi-variate regression of the dependent variable on the ground deposition.

C.2 Regressions with grid-level fixed effects

In Section 7.1, we present the estimation results from regressions with fixed effects at a $120 \times 120km$ grid-cell level, whereby we repeatedly estimate the same specification but randomly change the locus of the grid in every replication. Table 10 reports additional results of this exercise. Column (1) reports the average point estimate from 500 replications. In most cases, the estimates are smaller than the baseline results, which is due to the difference in fixed effects. Column (2) reports the average p-values. The statistical significance found here corresponds to the one of our baseline results. While the p-values are higher than those in Table 3, even in this more restrictive specification, the effects for math and reading are significant at the 10%-level and the index for crystallized intelligence and listening are significant at the 5%-level. The same pattern is illustrated by Columns (3)-(5).

Table 9: Adjusted R^2 for main results.

	(1)	(2)	(3)	(4)
A. Individual test scores				
Math	-0.00	0.20	0.21	0.21
Reading	-0.00	0.19	0.21	0.21
Listening comprehension	0.00	0.11	0.12	0.12
ICT	-0.00	0.22	0.22	0.22
Scientific literacy	-0.00	0.20	0.20	0.20
Reasoning	-0.00	0.13	0.14	0.14
Reading speed	-0.00	0.11	0.11	0.12
Perceptual speed	-0.00	0.27	0.27	0.27
B. Indices				
Cognitive skill index	-0.00	0.21	0.22	0.22
Crystallized intelligence index	-0.00	0.20	0.21	0.21
Fluid intelligence index	-0.00	0.19	0.20	0.20
Controls:				
Individual characteristics	No	Yes	Yes	Yes
Municipality characteristics	No	No	Yes	Yes
State FE	No	No	No	Yes

Notes: This table displays the adjusted R^2 for the baseline results presented in Columns (5)-(8) in Table 3.

Table 10: Estimation results, regressions with grid-cell fixed effects

	Average coefficient	Average p-value	Share of p-values with		
			$p < 0.1$	$p < 0.05$	$p < 0.01$
	(1)	(2)	(3)	(4)	(5)
A. Individual test scores					
Math	-0.007	0.087	0.678	0.398	0.040
Reading	-0.009	0.073	0.800	0.322	0.000
Listening comprehension	-0.009	0.036	0.966	0.746	0.120
ICT	-0.005	0.213	0.016	0.000	0.000
Scientific literacy	-0.004	0.309	0.014	0.000	0.000
Reasoning	0.003	0.647	0.000	0.000	0.000
Reading speed	-0.001	0.681	0.018	0.002	0.000
B. Indices					
Cognitive skill index	-0.004	0.102	0.540	0.228	0.014
Crystallized intelligence index	-0.007	0.048	0.936	0.616	0.016
Fluid intelligence index	-0.002	0.641	0.000	0.000	0.000
Controls:					
Individual characteristics	Yes				
Municipality characteristics	Yes				
State FE	Yes				

Notes: This table displays the average point estimates and p-values from 500 replications of a regression of the outcome listed on the left on the ground deposition of Cs137 with grid-cell fixed effects. In each replication, a $120 \times 120km$ grid has been randomly shifted north-south and east-west. Columns (3)-(5) report the shares of estimates with p-values smaller than the thresholds stated in the headings. In all regressions, we control for individual and municipality characteristics. Standard errors are clustered at county level.

D Testing for systematic sample selection

D.1 Selective mortality

Figure 6a shows how regional differences in CS137 kBq/m^2 affects the mortality rate of the NEPS age cohort on county level between 1995 and 2010. We use this as a proxy for health related selection into our sample. However, even when splitting this sample into narrower age cohorts like 6b we do not find any affects.

$$m_{crst} = \alpha + \rho_{rt}\ Cs137_{crs} + \boldsymbol{X'_{cs}}\boldsymbol{\kappa} + \delta_s \varepsilon_{crst}. \tag{5}$$

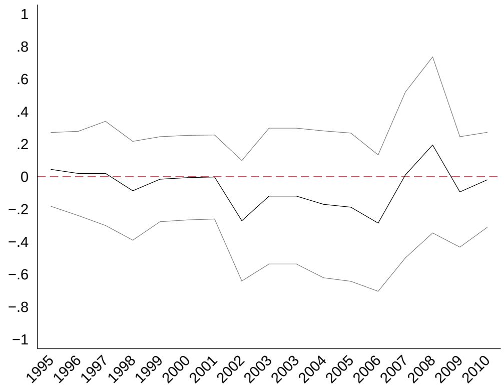
The number of deaths m_{cst} of age cohort r in county c state s in year t is regressed on the level of ground deposition of Cs137 in May 1986 in the same county. To obtain the level of ground deposition for each county, we match the radiation data based on the county centroid. The vector of controls, $\boldsymbol{X_{cs}}$, includes county characteristics, namely the level of rainfall altitude at the centroid and the total population in the country. In addition, we control for state fixed effects. The error term ε_{cst} summarizes all determinants of mortality not captured by the regressors. The coefficient ρ_r measures the reduced-form effect of exposure to radiation in April 1986 on mortality between 1995 and 2010.

D.2 Design-based attrition

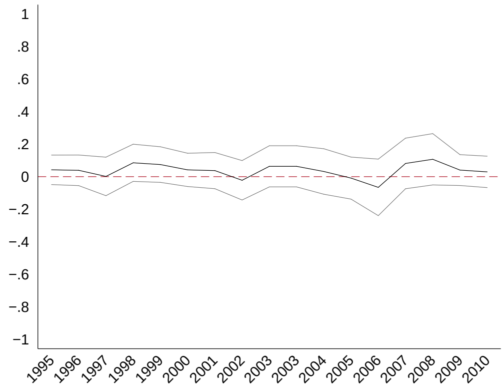
Table 11: Selection into competence tests

	(1)	(2)	(3)	(4)
Math	-0.002	-0.003	-0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.000)
Reading	-0.000	-0.001	0.000	-0.000
	(0.001)	(0.002)	(0.002)	(0.000)
Listening comprehension	-0.000	0.002	0.002*	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
ICT	0.002**	0.002	0.003**	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Scientific literacy	0.002**	0.002	0.003**	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Reasoning	-0.000	0.002	0.003*	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Reading speed	-0.001	-0.003	-0.001	-0.000
	(0.001)	(0.002)	(0.001)	(0.000)
Perceptual speed	-0.000	0.002	0.003*	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Controls:				
Municipality characteristics	No	Yes	Yes	Yes
State FE	No	No	Yes	Yes
Individual characteristics	No	No	No	Yes

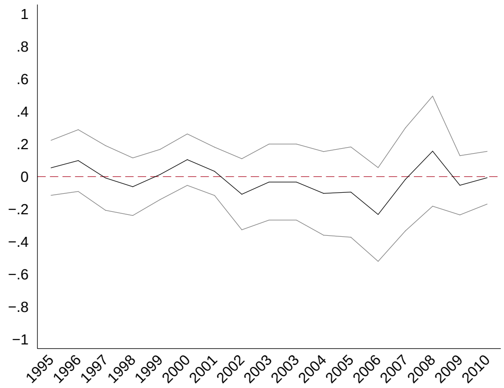
Notes: This table displays the results of separate regressions of dummy variables — indicating if an individual participated in the test or not — listed on the left on the ground deposition of Cs137, controlling for the variables indicated at the bottom. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.



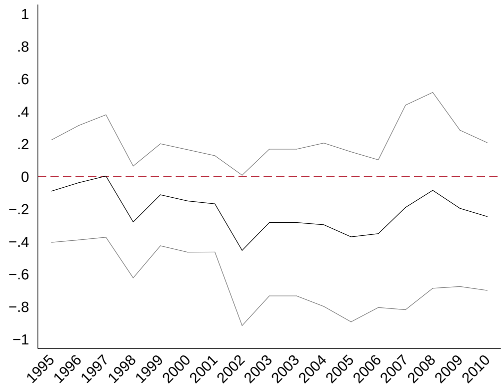
(a) All cohorts (1955-1985)



(b) Cohort 1975-1985



(c) Cohort 1965-1975



(d) Cohorts born before 1965

Figure 6: Radiation exposure and mortality.

Notes: This graph displays the estimated effect of radiation exposure on mortality in a given year. Both radiation and mortality vary at the county-level. In all regressions, we control for county-level characteristics as well as state fixed effects. The lines in each panel represent the point estimates and 95%-confidence intervals based on separate regressions for each year. Panel (a) presents the estimates of ρ_t for all cohorts in our estimation sample. Panels (b), (c), and (d) display the estimates of ρ_{rt} for distinct cohorts.

Table 12: Attrition

	(1)	(2)	(3)
A. Participation in competence test			
Cs137 kBq/m^2	-0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
(N)	5827	5827	5827
B. Missing personal information			
Cs137 kBq/m^2	0.001	0.000	-0.000
	(0.001)	(0.001)	(0.001)
(N)	4528	4528	4528
C. Municipality included in sample			
Cs137 kBq/m^2	0.0000	0.0004	-0.000
	(0.0002)	(0.0002)	(0.0002)
(N)	11197	11197	11197
Controls:			
Municipality characteristics	No	Yes	Yes
State FE	No	No	Yes

Notes: This table displays the results of regressions of indicators for participation or attrition on the level of fallout in 1986. In all regressions, we control for municipality characteristics and state fixed effects. In Panel A, the dependent variable is a binary indicator that equals unity if a person participated in the competence test. In Panel B, the dependent variable equals unity if the person is excluded from the estimation sample due to missing personal information. In Panel C, the dependent variable is an indicator

D.3 The cognitive skills index with attrition

Table 13 shows how the effect of CS137 kBq/m^2 changes by the number of observation. Regressing eight tests or less on CS137 kBq/m^2 equals the cognitive skill index. Effect size and significance slightly decreases if we exclude those who performed all tests, decreasing the sample by about 1000 observations. Decreasing the sample about another 1000 observation, namely those who performed seven tests, decreases effect size again and lead to insignificant results.

Table 13: The cognitive skills index with different definitions

	(Coef.)	(N)
All eight tests	-0.014***	1029
	(0.005)	
At least seven tests	-0.012***	2151
	(0.004)	
At least six tests	-0.013***	2352
	(0.004)	
At least five tests	-0.010***	3197
	(0.004)	
At least four tests	-0.010***	3455
	(0.003)	
At least three tests	-0.010***	3929
	(0.003)	
At least two tests	-0.009***	4417
	(0.003)	
At least one test	-0.008***	4423
	(0.003)	
Controls:		
Individual characteristics	Yes	
Municipality characteristics	Yes	
State FE	Yes	

Notes: This table displays the results of regressions of the standardized cognitive skills index on the level of ground deposition of Cs137 and the controls listed at the bottom. In each row, we consider different sample definitions. In row one, the index is based on participants who completed all eight tests. In the second row, we consider all participants who completed at least seven tests, etc. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

E Inference

E.1 Randomization inference

To assess the reliability of the inference based on the results presented in Section 6, we perform permutation tests. At the core of this test is a placebo distribution of point estimates, that is, a distribution of estimates that would occur if the relationship between radiation and cognitive skills was complete noise. To obtain this distribution, we randomize the level of Cs137 and the cognitive skill index seperatly across observations and estimate the regression presented in Table 3 Columns (3) and (4) with the standardized index as dependent variable. We repeat this procedure 10,000 times.

Figure 7a displays the placebo distribution of 10,000 estimates with randomization across all observations, which allows us to assess the inference in a model without state fixed effects (Table 3 Column (3)). If the relationship was pure noise, a point estimate at least as extreme as -0.008 would be very unlikely to occur. In fact, in 10,000 replications, such a result did only occur once.

The distribution in figure 7b corresponds to the estimations with state fixed effects presented in Table 3 Column (4). In this test, we randomize the regressor within states and otherwise follow the same procedure as before. Again, an estimate at least as extreme as our point estimate of -0.008 would be very unlikely to occur by chance. In 10,000 estimations, it occurred 26 times, i.e. in 0.026% of all cases.

Figure 7c displays the placebo distribution of 10,000 estimates with randomization of the outcomes of cognitive skill index across all observations, which is another approach to assess the inference in a model without state fixed effects (Table 3 Column (3)). In 10,000 replications, such a result equal to the point estimate in table 3 in Column (3) did only occur 112 times.

The distribution in figure 7d corresponds to the estimations with state fixed effects presented in Table 3 Column (4). In this test, we randomize the regressor within states and otherwise follow the same procedure as in Panel (c). A point estimate of 0.008 would only occure in 138 of 10,000 cases.

E.2 Multiple hypothesis testing

Figure 14 displays p-values using false discovery rate control as supposed by (Anderson, 2008) to control for the expected pro- portion of rejections that are type I errors.

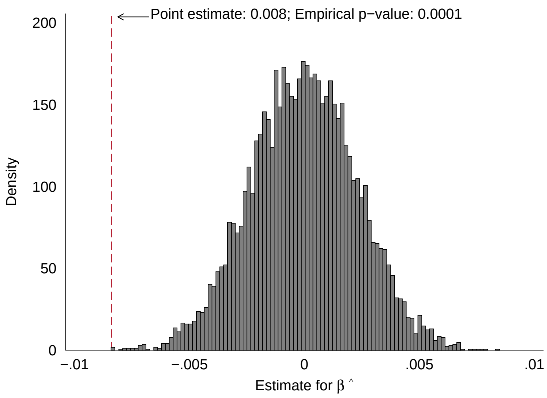
We first Compute the exact p-value for each test. Order the p-values from largest to smallest. For the first test (i=1) we adjust the p-value to be

$$\frac{pm}{m-(1-1)}$$

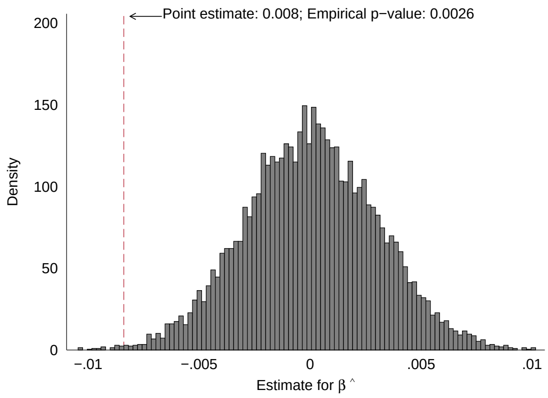
where m is the number of tests. For the following tests i, we adjust the p-value to be

$$\frac{pm}{m-(i-1)}$$

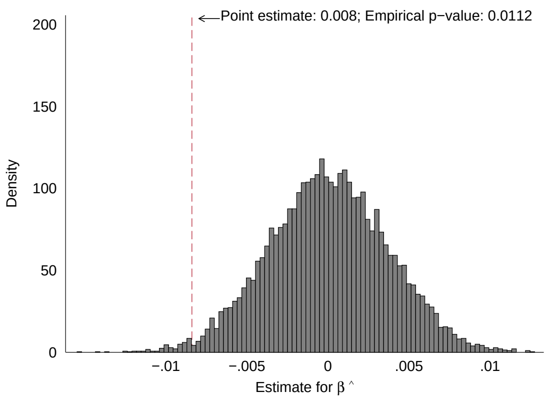
Table 15 shows inverse co-variance weighted indexes of those which we use in our analysis —flowing the approach of (Anderson, 2008) Based on the cognitive skill index, the crystallized intelligence index and the fluid intelligence index we create new indexes where we weighted their inputs by the sum of their row entries in the inverted co-variance matrix of each index leading to an efficient generalized squared estimator. This adjustment accounts for multiple hypothesis testing taking correlations among variables into account. Due to correlations strong correlations among test dimension, it is not surprising that the inverse co-variance weighted indexes only marginally differ.



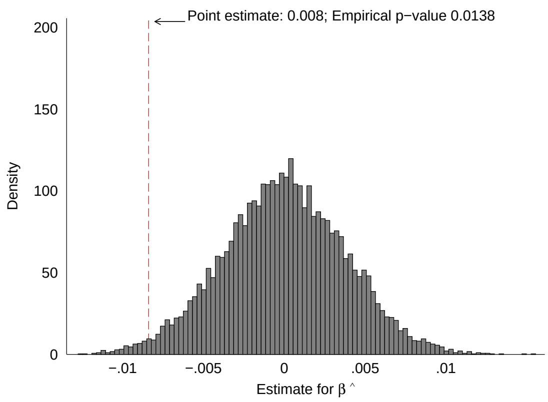
(a) [Full randomization od CS137



(b) [Randomization of CS137 within states



(c) [Full rando. of cog. skill index



(d) [Rando. of cog. skill index within states

Figure 7: Randomization inference

Notes: This figure displays the empirical distributions of the estimates for $\hat{\beta}$ under the null hypothesis of no treatment effect based on 10.000 replications. In each replication, we randomize the ground deposition while keeping the outcome — the standardized index — and all other regressors fixed. In Panel (a), the treatment is randomized across all observations; in Panel (b), it is randomized across observations within states. The vertical lines indicate the point estimate reported in Table 3 as well as the empirical p-values based on the empirical distributions.

Table 14: Q-values

	(1) p-values	(2) q-values
Math	0.001	0.005
Reading	0.004	0.018
Listening	0.017	0.045
ICT	0.092	0.148
Science	0.179	0.206
Reasoning	0.821	0.821
Reading Speed	0.023	0.046
Prectional Speed	0.176	0.206

Notes: This table displays the conventional p-values (Column (2)) as well as the p-values adjusted for multiple hypothesis testing (also called q-values, Column (1)). The p-values in Column (2) are based on standard errors clustered at the county-level.

Table 15: Summary index tests

	(1) Unweighted	(2) Weighted
Cognitive skill index	-0.008***	-0.007**
	(0.003)	(0.003)
Crystallized intelligence index	-0.009***	-0.009***
	(0.003)	(0.003)
Fluid intelligence index	-0.006*	-0.006*
	(0.003)	(0.003)
Controls:		
Individual characteristics	Yes	Yes
Municipality characteristics	Yes	Yes
State FE	Yes	Yes

Notes: This table displays the results of regressions of the indices listed on the left on the ground deposition of Cs137 and the controls listed at the bottom. Column (1) reproduces the baseline results from Table 3 Column (8), whereby the standardized indices are unweighted.

F Geographic information

Figure 8b shows the distribution of 3.448 measurement points for soil contamination measured by a in-situ gamma ray spectrometer executed during summer and autumn 1986. The reference date is May 01 1986. Counting back radiation is possible due to the well-known decay time of cesium which is location-independent. Measurement points are assigned to their geographical coordinates (Degrees minutes, seconds). Most measurement points in Bavaria do not comprise seconds. On the basis of experiences with the measurement of fallout by nuclear weapons tests in the sixties a protocol already existed of how in-situ measurements had to be executed. Measruing point should be located at meadows with a minimum distance to sealed surfaces of 30 meter. The surface should be flat and as far away as possible from trees and bushes.

Due the federal structure of Germany, several institutions were involved in the collection of measurements (Bavarian State Ministry for Regional Development and Environmental Issues; The Bavarian State Ministry for Food, Agriculture and Forestry; The Institute for Water, Soil and Air Hygiene of the Federal Health Office; State Office for Environmental Protection in Baden-Wuerttemberg; RWTH Aachen University). However, the leading institute was the Institute of Radiation Hygiene (ISH) of the former the German Federal Health Office (BGA) which coordinated, collected and evaluated measurements. Differences in the density of measurement points reveal spatial differences in the sampling process.

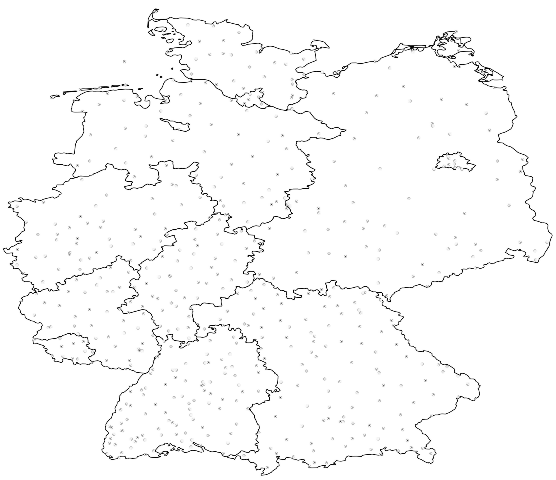
After the plume reached Germany, measurements were taken all over Germany. If high radiation was detected more measurements were taken in such a region. This explain clusters of measurements points and further explains the high density of measurement points in Bavaria. As Bavaria received the highest amount of fallout a measuring program was initiated with a 8x8 km grid (Winkelmann et al., 1986) (Winkelmann et al., 1989) (Fielitz and Richter, 2013).

In the GDR the "Staatliche Amt fuer Atomsicherheit und Strahlenschutz" (SAAS) was the only institute responsible for the execution and evaluation of measurements. A country-wide measurement program was initiated with a 8x8 km grid (Bundesamt für Strahlenschutz, 2016). However, figure 8b reveal that the measurement points in the GDR in our dataset is not as dense as in Bavaria. After the collapse of the GDR the Institute for Water, Soil and Air Hygiene of the Federal Health Office (WaBoLu) combined the data of in-situ gamma ray spectrometer collected by the GDR and the FRG to the dataset we are using. Only highly reliable measurements were used by the WaBoLu which explains missing measurements points in the GDR. In 1994 the WaBoLu was integrated in the Federal Environment Agency. The Federal Office for Radiation Protection provided us the radiation data which is the successor organization of the (ISH).

Figure 8b shows the distribution of 544 weather stations provided by the German Meteorological Service. In the FRG, the German Meteorological Service run these station. The stations in the GDR were operated by the Meteorological Service of the GDR which was integrated in the German Meteorological Service. In comparison to figure 8a a uniform distribution is evident across the county. The principal aim of this distribution is the collection of weather data which is representative for the whole county. Furthermore location requirements determine the exact distribution of weather stations. The inclination of the surrounding terrain, for example, should not exceed a specific limit, operation near high buildings is not possible and measurement operation should be executable for at least 10 years (Wetterdienst, 2017).

The amount of precipitation Germany received between April 30 and May 08, 1986 is shown in Figure 10a. Darker color represents higher precipitation. We determine this period as critical period based on our observations in 1. Comparing this figure with in **??** reveals a high correlation between CS137 and precipitation. However, high level of precipitation does only result in high exposure if radioactive plume was present. Figure 10b shows the average yearly precipitation between 1981 and 1985. Comparing figure 10a and 10b reveal the random process of rainfall given a short observation period. Some regions which normally receive high level of precipitin stayed relatively dry within the critical period whereas other regions which tend to receive less perceptions faced heavy rainfalls. However, there are regions which tend to receive more precipitation in the long-run also receive more precipitation in the short-run. One main factor for spatial differences in precipitation is altitude due to orographic rainfall. Figure **??** shows a rough distribution of mountainous regions in Germany where darker color means higher altitude. We generated this figure with the data on the altitude of each municipality center provided by the Federal Agency for Cartography and Geodesy. Southern parts of Germany — which received most of the fallout — tend to be more mountainous. Altitude could be an confounding factor if it affects population density. Therefore, we show population density in figure **??**. Darker color means higher population. Again, this figure was generated with the municipality data provided by the Federal Agency for Cartography and Geodesy, showing the population of each municipality in 1997. Comparing **??** and **??** reveals that regions at higher altitude tend to be less populated.

(a) CS137



(b) Precipitation

Figure 8: Measurement points. Source: The German Meteorological Service, The Federal Office for Radiation Protection

Table 16: Different data linkage

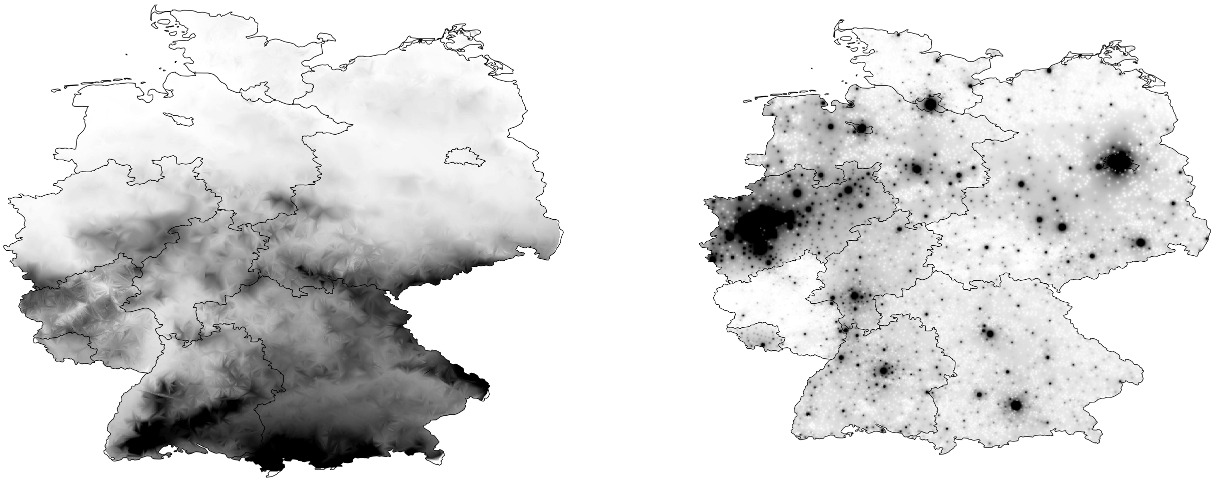
	(c4)	(cc)	(b4)	(bc)	(m4)	(mc)	(avg)	(avgw)
A. Individual test scores								
Math	-0.011***	-0.008***	-0.012***	-0.009***	-0.014***	-0.011***	-0.013***	-0.012***
	(0.003)	(0.002)	(0.005)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)
Reading	-0.014***	-0.008**	-0.018***	-0.011**	-0.017***	-0.008*	-0.016***	-0.015***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Listening comprehension	-0.009**	-0.005*	-0.010**	-0.006**	-0.008	-0.005	-0.012***	-0.012**
	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.004)
ICT	-0.006*	-0.002	-0.008*	-0.004	-0.003	0.001	-0.006*	-0.006*
	(0.003)	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Scientific literacy	-0.004	-0.002	-0.006	-0.004	-0.003	-0.002	-0.005	-0.005
	(0.003)	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Reasoning	-0.001	0.001	0.001	0.002	0.002	0.001	-0.002	-0.002
	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Reading speed	-0.008**	-0.005*	-0.011***	-0.005	-0.012**	-0.004	-0.011**	-0.010**
	(0.003)	(0.002)	(0.04)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
Perceptual speed	-0.004	0.002	-0.006	-0.003	-0.005	-0.002	-0.006	-0.006
	(0.003)	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
B. Indices								
Cognitive skill index	-0.008***	-0.005**	-0.011***	-0.006**	-0.009**	-0.004	-0.010***	-0.010***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Crystallized intelligence index	0.009***	-0.005**	-0.012***	-0.007***	-0.010**	-0.005	-0.011***	-0.010***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Fluid intelligence index	–0.006*	-0.003	-0.007*	-0.002	-0.006	-0.003	-0.008*	-0.007*
	(0.003)	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Controls:								
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table displays the main estimation results in the first column (c4), equal to column 8 in Table 3 and seven additional columns with varying dependent variables due to different data linkage procedures of caesium-137. Each coefficient is the result of a separate regression of the outcomes listed on the left on the ground deposition of Cs137 in kBq/m^2, which varries by data linkage ((c4) link via centriod, distance weigthed, 4 measuring points; (cc) link via centriod, closest measuring point; (b4) link via balancing point, distance weighted, 4 measuring points; ((bc) link via balancing point, closest measuring point; (m4) link via mode, distance weighted, 4 measuring points; (mc) link via mode, closest measuring point; (avg) average radiation within a municipality; (avgw) population weighted average radiation within a municipality, controlling for the variables indicated below. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table 17: Wild bootstraped se's

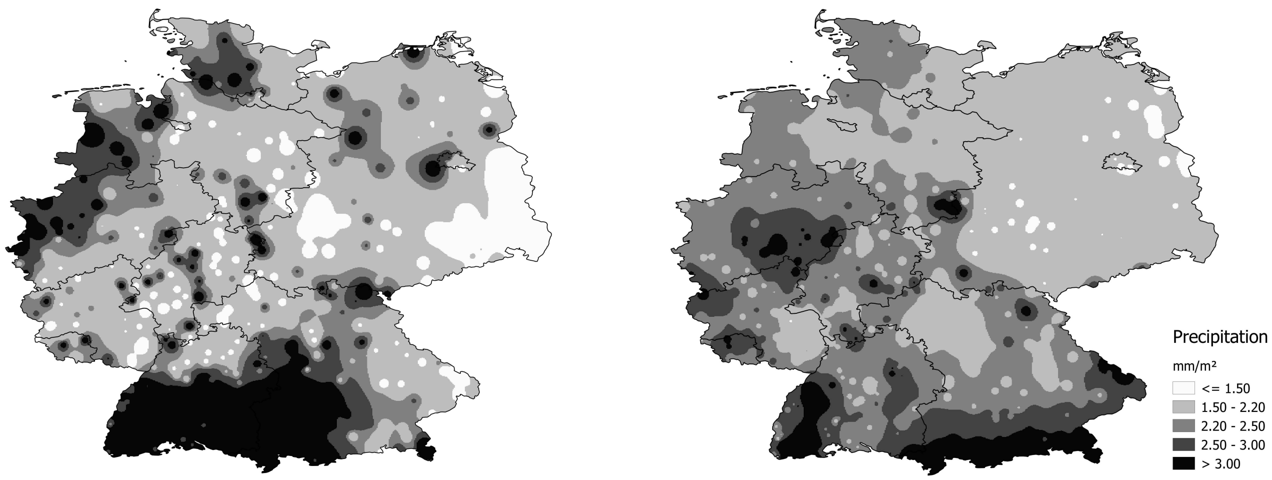
	(1)	(2)	(3)	(4)
A. Individual test scores				
Math	0.003	-0.001	-0.012	-0.011*
	(0.004)	(0.003)	(0.008)	(0.006)
Reading	-0.002	-0.005	-0.014**	-0.014*
	(0.009)	(0.003)	(0.006)	(0.008)
Listening comprehension	-0.003	-0.006	-0.009	-0.009
	(0.004)	(0.004)	(0.006)	(0.007)
ICT	0.000	-0.003**	-0.005*	-0.006**
	(0.008)	(0.001)	(0.003)	(0.003)
Scientific literacy	0.001	-0.000	-0.004*	-0.004
	(0.002)	(0.003)	(0.002)	(0.003)
Reasoning	0.002	-0.003*	-0.002	-0.001
	(0.004)	(0.002)	(0.005)	(0.002)
Reading speed	-0.001	-0.006	-0.011	-0.008*
	(0.016)	(0.005)	(0.007)	(0.005)
Perceptual speed	0.003	0.000	-0.004**	-0.004**
	(0.003)	(0.001)	(0.002)	(0.002)
B. Indices				
Cognitive skill index	0.000	-0.003	-0.009**	-0.008*
	(0.007)	(0.002)	(0.004)	(0.004)
Crystallized intelligence index	0.000	-0.003	-0.008**	-0.009*
	(0.000)	(0.003)	(0.004)	(0.005)
Fluid intelligence index	0.001	-0.003	-0.007	-0.006
	(0.013)	(0.002)	(0.006)	(0.04)
Controls:				
Individual characteristics	No	Yes	Yes	Yes
Municipality characteristics	No	No	Yes	Yes
State FE	No	No	No	Yes

Notes: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.



(a) Topography (b) Population density

Figure 9: Altitutde and population density, darker means higher. Source: Federal Agency for Cartography and Geodesy



(a) Critical 10 days in 1986 (b) Between 1981 and 1985

Figure 10: Average daily Precipitation. Source: The German Meteorological Service

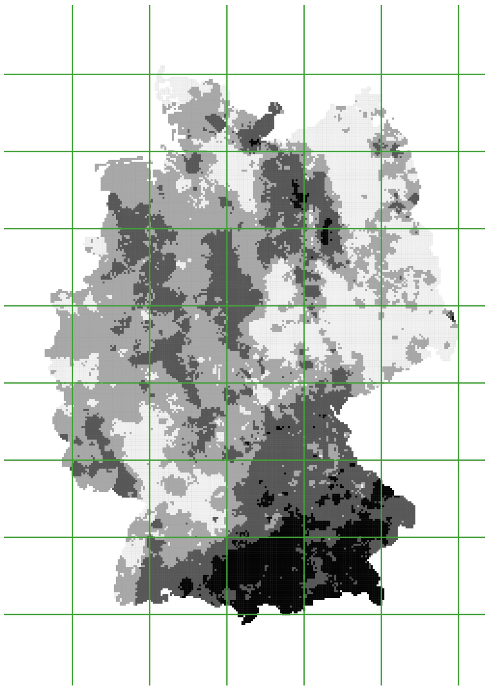


Figure 11: 120km × 120km grid cells and regional variation in caesium-137 ground contamination in May 1986

G Control variables