The Cognitive Meltdown: Radiation and Human Capital after Birth

Benjamin Elsner
University College Dublin, IZA, and CReAM

Florian Wozny
IZA

March 27, 2019

Abstract:
This paper studies the long-term effect of post-natal radiation exposure on cognitive skills. We use regional variation in nuclear fallout after the Chernobyl disaster in 1986, which led to an increase in radiation levels in most of Europe. In order to identify a causal effect, we exploit the fact that the degree of soil contamination depended on rainfall within a critical ten-day window after the disaster. Based on unique geo-coded survey data from Germany, we show that people who lived in highly-contaminated areas in 1986 perform significantly worse in standardized cognitive tests 25 years later. This effect is driven by the older cohorts in our sample, whereas we find no effect for people who were first exposed at the ages of 0-7. These results suggest that pollution can have adverse effects even when people are first exposed as adults, and point to significant external costs of man-made sources of radiation.

JEL-Classification: J24, Q53

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

We would like to thank Hoyt Bleakley, Leonardo Bursztyn, Deborah Cobb-Clark, Tommaso Colussi, Joan Costa-i-Font, Thomas DeLeire, Olivier Deschenes, Paul Devereux, Jon Guryan, Adam Isen, Ingo Isphording, Shelly Lundberg, Keith Meyers, Nico Pestel and Erik Plug, as well as audiences at numerous conferences and seminars for helpful comments. This paper has been previously circulated under the title *The Human Capital Cost of Radiation: Long-Run Evidence from Exposure Outside the Womb.* We are grateful to the team of NEPS, in particular Tobias Koberg, for their invaluable help with the data and their patience. Nicolas Zimmer and Filippo Ricordi provided outstanding research assistance.

Corresponding author: Benjamin Elsner, University College Dublin, IZA, and CReAM. University College Dublin, School of Economics, Belfield, Dublin 4, Ireland. Phone, +353 1 716 8446, E-mail benjamin.elsner@ucd.ie

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, the 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

The last 40 years have seen a drastic increase in radiation exposure. Today, the average person in Europe and America receives about twice the annual dose of radiation compared with in 1980.[1] This increase is almost entirely due to man-made sources of radiation, namely medical procedures and nuclear power. CT scans, x-rays or mammograms expose patients to low doses of radiation. Moreover, many people were exposed to the nuclear disasters in Chernobyl and Fukushima, whose fallout has been widely distributed. Medical research shows that subclinical radiation damages human cells, which has potential knock-on effects on health and cognition. These effects may occur at all ages. They are particularly strong during critical periods — for example, during gestation or during puberty — although the biological effects are not limited to those periods. The existing literature has largely focused on the effect of in-utero exposure, documenting significant adverse effects of radiation exposure during pregnancy on education and labor market outcomes many years later (Almond et al., 2009; Heiervang et al., 2010; Black et al., 2013). However, there is little evidence on the long-term effects of radiation exposure after birth.

In this paper, we exploit a natural experiment to study the effect of post-natal exposure on cognitive test scores 25 years later. We use the fallout from the Chernobyl disaster in 1986 as a source of exogenous variation in radiation exposure, and focus on Germany, which received a large share of the overall fallout. The level of fallout in an area depended on rainfall levels while the plume was over Germany. In Munich, which saw heavy rainfalls during that period, the ground contamination was seven times higher than in Hamburg, where it rained very little. Due to its long half-life, the fallout led to a quasi-permanent increase in radiation levels. People in highly-contaminated areas were constantly exposed to higher radiation for the next 25 years.

Our study is based on the National Educational Panel Study (NEPS), a representative survey of the German population born between 1956 and 1986. NEPS offers three features that are key to our analysis. First, unlike most datasets used in the literature, it includes reliable measures of cognitive skills long after school-leaving age. This enables us to analyze the impact of radiation on cognitive skills up until a person's late-fifties. Second, at the time of Chernobyl, around half of the sample were over 18 years old, allowing us to study the long-term effect of radiation for people who were first exposed as adults. Finally, the survey contains a detailed residential history for each participant, allowing us to link personal information with data on radiation in the the respondent's place of residence in 1986. We match the NEPS with fine-grained decay-corrected radiation data from a measurement program rolled out by the German government between 1986 and 1989. The combination of the two datasets enables us to run a reduced-form regression of cognitive skills in 2010 and afterwards on the initial level of fallout in a person's place of residence in 1986.

In order to identify a causal effect, we exploit the fact that the fallout was determined

[1] See National Council on Radiation Protection and Measurements (2009).

by rainfall during a critical period of ten days in early May 1986. Balancing tests show that — conditional on state fixed effects — the amount of fallout in a respondent's municipality of residence is uncorrelated with observable characteristics. This suggests that exposure is unrelated with residential sorting and supports the identifying assumption that — within states — the fallout was as good as randomly assigned. In order to further strengthen our identification, we apply an instrumental variable strategy that leverages rainfall patterns in May 1986 as well as the movement of the radioactive plume. Following an approach from physics, our instrument predicts the local amount of fallout from the interaction of the abnormal level of rainfall — the deviation from the average rainfall early May — and the amount of radioactive matter in the plume. Because large amounts of fallout were rained down when the plume entered Germany in the south-east, less fallout was available as the plume moved north-west. The identifying assumption behind this strategy is that abnormal rainfall levels during ten days in May 1986 have no effect on cognitive skills in 2010 other than through fallout. This assumption is supported by balancing and placebo tests.

This study delivers two central findings. First, we show that people exposed to higher radiation from 1986 onwards have significantly lower cognitive test scores 25 years later. A one-standard-deviation higher initial exposure in 1986 reduces test scores by between 4.7 and 8.2% of a standard deviation. Second, we document differences in the effect across age cohorts. While we find strong negative effects for older cohorts — born between 1956 and 1979 — we find no effect for younger cohorts who were first exposed as children. Upon first glance, this result seems at odds with the common finding that pollution matters most when people are exposed in the womb or during early childhood (Almond and Currie, 2011; Graff Zivin and Neidell, 2013). There are several potential explanations for this finding. First, our study identifies a different biological channel. Almond et al. (2009) and Black et al. (2013) identify the effect of radiation during neurogenesis — the development of the central nervous system in a fetus — which occurs between weeks 8 and 25 of gestation. By contrast, our study focuses on the impact of exposure after birth. At this stage, radiation affects the human organism by inducing a stochastic error in the reproduction of cells, which can lead to cognitive decline during older ages (Rola et al., 2004). Because this effect only becomes noticeable during older age, the younger cohorts may simply be too young to experience negative effects today. To further corroborate this mechanism, we show that areas with greater levels of fallout have a higher incidence of dementia almost 30 years later. An alternative explanation is that the younger cohorts — or their parents — engaged in compensating behavior to avoid radiation exposure. While there is anecdotal evidence of such behavior immediately after Chernobyl, we discuss why it is hardly plausible that people could avoid exposure over a 25-year period.

Our results imply that post-natal exposure to radiation has stronger negative effects than previously thought. Over 25 years, people who lived in an area with a one-standard-deviation higher level of fallout received a cumulative dose equivalent to two mammograms or 40% of a CT scan. According to our estimates, this dose reduces cognitive test scores by the equivalent of

0.05-0.08 school years. These effects are similar to those found for cancer patients who underwent radiotherapy (Hall et al., 2004; Pearce et al., 2012), although they are an order of magnitude smaller than the effect of in-utero exposure. For example, Almond et al. (2009) find an effect that is nine times larger. However, it is important to consider the number of people exposed in Germany at different ages. At the time of Chernobyl, around 200,000 fetuses were in the womb during the critical period between 8 and 25 weeks of pregnancy. The birth cohorts in our sample — 1956 to 1985 — amounted to 24 million people. Therefore, although the post-natal effects are smaller, they are economically significant because they apply to much larger parts of the population.

This paper contributes to two strands of literature. First, it fills a gap in the literature on the effect of pollution on human capital. This literature has mainly focused on two effects, namely i) the long-run effect of in-utero exposure on outcomes such as birth weight, crime rates, wages or IQ (Almond and Currie, 2011; Graff Zivin and Neidell, 2013) and ii) the contemporaneous effect of exposure during adulthood on productivity, often measured on the same day.[2] However, there is little evidence on the long-run effect exposure to pollution *after* birth. By showing that constant exposure to radiation after birth has affected cognitive skills over more than two decades, we are among the first papers to document such an effect.

Second, this paper provides new evidence on the external cost of nuclear power. While not occurring often, nuclear disasters such as Chernobyl and Fukushima can be highly destructive at the epicenter and spread large amounts of fallout across the globe. Besides the aforementioned studies on exposure *in utero* (Almond et al., 2009; Heiervang et al., 2010), several studies suggest that Chernobyl increased the incidence of cancer (Nature, 1992; Auvinen et al., 2014; Alinaghizadeh et al., 2016), although other studies find no significant effect (Rumyantsev et al., 2011). Moreover, Lehmann and Wadsworth (2011) and Danzer and Danzer (2016) document negative effects of the fallout on wages and subjective well-being in Ukraine, the country in which Chernobyl is located. By focusing on a country located over 1000km away from Chernobyl, we complement this literature by showing that nuclear power imposes an externality over long distances.

2 The Chernobyl fallout in Germany

The Chernobyl nuclear disaster The Chernobyl nuclear disaster in 1986 is one of the two largest nuclear accidents in history. 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 reactor. In the two weeks following the accident, several trillion Becquerel of radioactive matter were emitted from the reactor, stirred up into the

[2]See, for example, Currie et al. (2009) for the effect of pollution on school absences, Ebenstein et al. (2016) for the effect of test scores, as well as Graff Zivin and Neidell (2012) and Lichter et al. (2017) for the effect on productivity. Exceptions in this literature are Currie and Neidell (2005) and Arceo et al. (2016), who find short-term effects of air pollution on the mortality of young children.

atmosphere, and — through strong east winds — carried all over Europe.[3] The most affected countries were Belarus, Ukraine as well as the European part of Russia, although other regions, such as Scandinavia, the Balkans, Austria and Germany also received considerable amounts of fallout. The only other 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 south-east and made its way north-west before disappearing over the North Sea on May 8. The fallout comprises 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.[4] Among the four isotopes, soil-bounded Cs137 is today considered the only relevant source of radiation in Germany that can be ascribed to the Chernobyl disaster (Hachenberger et al., 2017). From 1986 to 1989, the governments of West and East Germany rolled out a comprehensive program to measure radiation across the country. At over 3,000 temporary measuring points, gamma spectrometers measured the radiation of Cs137. Based on the decay of the isotopes, all measurements were backdated to May 1986.

The deposition of the fallout varies considerably across regions, and depends on the amount of rainfall within a critical time window. Regions with heavy rainfall while the radioactive plume was hanging over Europe received large amounts of fallout whereas regions without rainfall received little to none. Figure 1a displays the ground deposition of Cs137 in May 1986. Because Cs137 rarely occurred in Germany before 1986, the displayed variation is almost entirely due to 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). The majority of the population 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.[5]

For affected regions, the nuclear fallout represented a quasi-permanent 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. Therefore, a person who has been living for the last 25 years in a highly-affected area has been constantly exposed to a higher dose of radiation than someone living the entire time in a less affected area. In 2010, the first year in which we measure

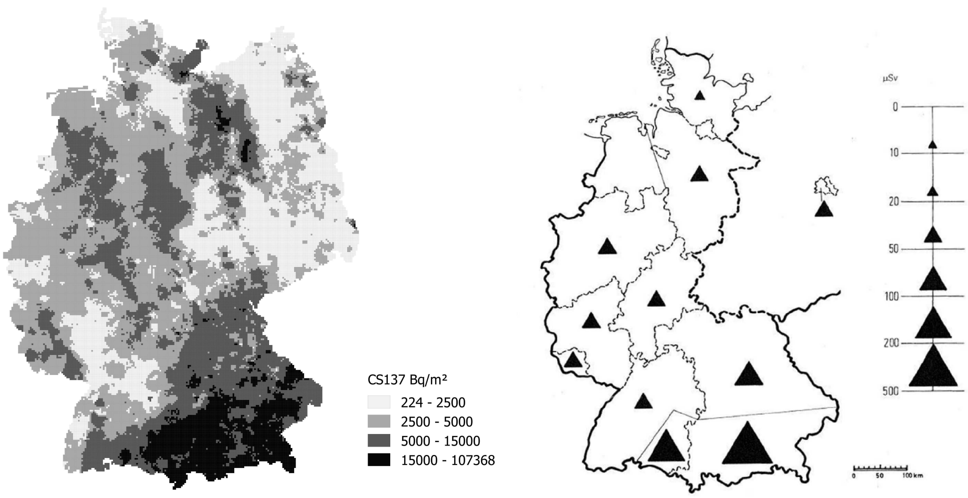
[3]Becquerel (Bq) is a unit of radioactivity. One Bq defines the activity of radioactive material in which one nucleus decays per second. In the following, we use kilobecquerel (kBq). One kBq equals 1000Bq.

[4]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.

[5]See Figure 5b in Appendix A.5 for the distribution of ground deposition in the German population.

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.

The German Agency for Radiation Protection (BfS) estimates that the cumulative effective radiation dose induced by Chernobyl between 1986 and 2010 was 0.6mSv. This amounts to 30% of the annual effective dose the average German receives from natural background radiation in one year (2mSv), or the dose received during 30 chest x-rays. However, the effective dose from Chernobyl varied considerably across regions. In Munich, one of the most affected cities, the cumulative effective dose over 25 years was 2.1mSv.[6] Due to the decay of the radioactive matter, the annual effective dose declined over time. The BfS estimates that the dose in the first year — when radioactive particles were in the air and, in general, the radiation was highest — accounted for 21% of the cumulative dose over 25 years. In 1987, the dose accounted for 11%, and it has been declining at an annual rate of 4% since.



(a) Ground deposition of CS137 (b) Information released in 1986

Figure 1: 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.)

Information about the nuclear disaster in the German public The German public learned about the nuclear accident several days after it occurred, and — in most parts of

[6] 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 unrelated to the length of low-dose exposure (Leuraud et al., 2015). However, it should be noted that the average exposure published by Bundesregierung (1986-1991) is more uncertain and is based on assumptions about daily activities, diet, etc.

the country — after the radioactive rain had fallen. Indications of a nuclear accident were first noticed in Sweden, where scientists measured abnormally high levels of radioactivity at the Forsmark nuclear power plant. The Soviet Union initially released no information about the accident, and its government only acknowledged it 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 were measured in several parts of the country, the government of the Federal Republic of Germany (FRG) introduced radiation limits on foods and warned the population of the consumption of dairy produce, 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 did not present a health hazard to the population. The information policy differed considerably between the 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 1b shows a map released by the German-Swiss Association 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 1a 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 —, it appears that these changes were short-lived. For example, Renn (1990) 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

In this section, we explain how radiation affects the body and why it is plausible that radiation exposure — even as an adult — can negatively affect health and cognition. Based on findings from the literature in radiobiology and medicine, we develop a simple conceptual framework that guides our empirical analysis.

3.1 Exposure to radiation

Humans can be exposed to radiation in three ways, namely through inhaling radioactive particles, ingesting contaminated foods, as well as external exposure, whereby radiation affects the body if a person 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 — and therefore strongly correlated with — a person's place of residence (Clark and Smith, 1988). By contrast, exposure through food may not necessarily result from contamination in the same locality, given that the food might 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 and diet. For example, people who spend more time outdoors are more exposed to cosmic radiation than those who spend most of their time indoors, while people who are physically active — and therefore breathe more — have a higher exposure through inhalation.

3.2 Impact on the human body

Radiation affects the human body through ionization, a process that damages the DNA and can lead to the dysfunction or death of cells. When it collides and reacts with the DNA in a cell, radiation can directly or indirectly damage the DNA. Radiobiology theory posits that a marginal increase in radioactivity linearly increases the probability that a cell is hit by an electron. A linear relationship emerges because during ionization the release of electrons follows a random process, whereby 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 (Brenner et al., 2003). The human organism has the capacity to repair damaged DNA. However, 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, the more likely that a critical mass of dysfunctional cells affects the functioning of organs and therefore leads to adverse health effects.

Impact on health The medical literature provides ample evidence of negative health effects. These effects can be either *deterministic*, whereby exposure to radiation almost inevitably affects a person's health, or *stochastic*, in which case radiation affects the likelihood of developing a health condition. Deterministic effects only result from high doses of radiation such as those encountered by survivors of the Hiroshima nuclear bomb or soldiers who cleaned up the nuclear waste in Chernobyl. By contrast, a low dose of radiation — defined as a short-term dose below 100 mSv — only induces stochastic health effects. At such levels, an increase in the dose raises the probability that a person experiences health problems later in life, but does not lead to the immediate dysfunction of organs (OECD, 2016). The medical literature provides evidence of the existence of stochastic health effects such as heart disease, stroke, digestive diseases, and respiratory diseases (Preston et al., 2003). People in at-risk occupations — for example, 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 cancer risk (UNSCEAR, 2008).

Impact on cognition — pre- and post-natal exposure The effect of radiation on cognitive and neuro-developmental functioning is an active research area in the sciences (OECD, 2016). The literature distinguishes between two types of effects, namely the effect of exposure during critical periods *in utero*, and after birth. It is well established that *in-utero* exposure can cause severe damage to the human brain. During weeks 8-15 after conception, the cerebral cortex is developed, which plays a key role in a person's cognition, language, perception, and consciousness. During this period, the brain of a fetus is particularly vulnerable. Empirical evidence shows that *in-utero* exposure to radiation leads to lasting cognitive impediments (Otake and Schull, 1998; Almond et al., 2009; Black et al., 2013).

Although the impact of radiation on cognition is strongest during pregnancy, it can unfold throughout a person's life. Recent research shows that radiation affects cell regeneration 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 exposure to radiation slows down the regeneration of brain cells, which in turn impairs cognitive performance. People who have been exposed to low-dose radiation during medical treatments are more likely to suffer from cognitive impairments several months to years later (Hall et al., 2004; Douw et al., 2009; Monje and Dietrich, 2012) and have a higher risk of developing dementia (Greene-Schloesser and Robbins, 2012). Therefore, it is biologically plausible that radiation affects cognition even if the exposure occurs long after birth.[7]

3.3 Conceptual framework and predictions

The scientific literature highlights two channels through which radiation exposure affects cognitive skills, namely its impact on brain cells as well as 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,

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

A cognitive test score y is produced with three inputs: a person's intelligence I, a person's health H, as well as any choices that people make in response to being exposed, summarized by $B = B(R)$. We think of B as compensating behaviors aimed at limiting or counteracting the impact of radiation. There are many possible behavioral responses; for example, investment

[7] Further evidence on the biological plausibility comes from experiments with mice. Rola et al. (2004) document lower brain activity among mice exposed to higher radiation. Kempf et al. (2016) show that mice who were exposed develop symptoms similar to Alzheimer's.

in education, moving to a less contaminated area, or changes in one's diet or exercise habits. We allow these behaviors to have a direct effect on test scores as well as an indirect effect by affecting people's intelligence or health.

Total differentiation of Equation (1) yields the proportional change of a test score in response to a change in levels of radiation,

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

The first two terms represent direct effects of radiation on intelligence and health ($\partial I/\partial R$ and $\partial H/\partial R$), combined with the impact of intelligence and health on test scores ($\partial F/\partial I$ and $\partial F/\partial H$). The remaining three terms represent the direct effect of behavioral responses on test scores ($\frac{\partial F}{\partial B}\frac{\partial B}{\partial R}$) as well as the indirect effects of behavioral responses that operate through intelligence and health.

Equation (2) allows us to generate hypotheses about the sign of each channel, although the sign of the overall effect remains ambiguous. In light of the existing evidence, the direct effects are either negative or zero. Low-dose radiation may have a negative impact on health and cognition unless the dose is too small to have any impact at all. On the other hand, the sign of the behavioral responses is either positive or zero. There is plenty of anecdotal evidence of behavioral responses to the Chernobyl disaster. According to a German survey from 1987, many people initially followed the government's recommendations to avoid certain foods and keep their children inside during the weeks after the radioactive rainfall (Peters et al., 1987). In addition, a study from Austria by Halla and Zweimüller (2014) shows that families responded to the fallout by having fewer children and reducing mothers' labor supply. However, as shown by Renn (1990), many behavioral responses — especially changes in diet and exercise habits — were fairly short-lived.

A further prediction is that both direct effects ($\partial I/\partial R$ and $\partial H/\partial R$) — and therefore the total effect — differ by age group. The replacement and repair of damaged cells is prone to a stochastic error that increases with age (UNSCEAR, 1994). For this reason, we expect the impact of radioactivity to be stronger among older rather than among younger people. Moreover, because within the brain radiation mainly affects the hippocampus, we would expect a stronger effect on skills based on crystallized intelligence than fluid intelligence, which is governed by a part of the brain with static cells.

Equation (2) also helps to interpret the estimates. Our regression — assuming that radiation exposure is exogenous — allows us to identify the *total* effect of radiation exposure on test scores, which comprises direct effects as well as compensating behaviors. If one was interested in 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 absent. In our analysis, while we cannot 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 Statistics

We link rich individual-level survey data with geo-coded information on radiation in a person's municipality of residence in May 1986. 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 addition, 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 NEPS, a rich representative dataset on educational trajectories in Germany. NEPS offers two features that are key to our analysis. First, the survey includes standardized competence tests that allow us to measure cognitive skills along various dimensions for people who aged between 24 and 58 years old in 2010. This represents a significant advantage over most datasets that include information on cognitive skills — notably the Scandinavian population register data — which typically only measure skills at school-leaving age (i.e. 18 or 19). Second, the NEPS includes detailed information on residential histories. For each respondent, it provides monthly spell data on their municipality of residence since their birth, allowing us to link personal characteristics and cognitive test scores measured after 2009 with data on radiation levels in the person's municipality of residence in May 1986.

The NEPS is supervised and hosted by the Leibniz Institute for Educational Trajectories (LIfBi, Blossfeld et al. (2011)). It comprises six starting cohorts, ranging from newborns to adults, which have been followed in multiple waves since 2010. In this study, we use the adult cohort of the NEPS (Starting Cohort 6 — SC6). More specifically, we use the so-called ALWA subsample of the adult cohort, which samples respondents born between 1956 and 1986. To set up the NEPS SC6, LIfBi took over a representative survey named Working and Learning in a Changing World (ALWA), which was conducted by the Institute for Employment Research (IAB) in 2007 with originally 10,404 respondents. The original aim of ALWA was to study geographic and occupational mobility, which is why IAB devoted considerable resources to eliciting residential and occupational histories. For further information on how this information was gathered, see Appendix A.1.

The NEPS SC6 includes all respondents of ALWA who were willing to enter the panel and be surveyed every year (N=8,997). Among those who agreed to be included, 6,572 actually participated.[8] A comparison of the ALWA subsample with the German Microcensus shows that

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

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.

4.2 Estimation sample

Our sample includes all survey participants who were born *before* Chernobyl. We exclude participants born after Chernobyl because the survey only sampled birth cohorts up to December 1986, leaving us with few participants who were born after Chernobyl. Moreover, because we are interested in the effect of post-natal exposure, excluding them ensures that our estimates are not confounded by exposure in utero, which operates through a different biological channel. Overall, we can link the municipality of residence in May 1986 for 5,844 participants. For the remaining 728 participants, we could not link the data due to missing municipality keys (402 obs.) or because they lived abroad in May 1986 (326 obs.). Observations with missing municipality keys include 140 participants born after April 1986.

To reduce classification error, we drop respondents who moved in May 1986 (34 obs.), for whom we cannot determine whether they moved before or after the radioactive plume reached Germany. We also drop all respondents who did not participate in the competence tests (1,265 obs.), as well as all participants for whom information on personal characteristics is missing (105 obs.). Our final estimation sample comprises 4,440 observations. 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 the residential history. Moreover, in order to address concerns about the representativeness of the estimation sample, we perform a series of balancing tests in Appendix B, which suggest that the missing information is unsystematic.

4.3 Cognitive tests

One of the core objectives of the NEPS SC6 was to collect data on the competencies of adults. The survey includes eight standardized cognitive tests that were modeled after well-established tests from psychology and related fields (Weinert et al., 2011). For our analysis, we use tests on *mathematical competence*, *reading competence*, *scientific literacy*, *listening comprehension*, *ICT literacy*, *reading speed*, *perceptual speed*, and *reasoning*. Appendix A.2 provides a detailed description of each test. In the empirical analysis, we use each test score as a separate outcome. In order to make the estimates comparable across outcomes, we standardize the test scores to a mean of zero and a standard deviation of one. Moreover, given that the test scores measure different aspects of the latent variable cognitive skills, we construct a standardized cognitive skills index that allows us to estimate the overall effect of radiation on the latent factor cognitive skills. To construct the index, we first sum over all eight standardized test scores, and then standardize this sum to a mean of zero and a standard deviation of one. Using the same standardization,

be contacted.

we construct sub-indices for skills governed by crystallized intelligence (math, reading, science, listening, ICT) and fluid intelligence (reading speed, perceptual speed, reasoning).

4.4 Municipality- and County-level data

Data on ground deposition Our regressor of interest is the ground deposition of Cs137 in kBq/m^2 in May 1986, which we use as proxy for Chernobyl-induced radiation in Germany. The regional concentration of Cs137 is strongly correlated with other Chernobyl-induced sources of radiation such as I131 or Sr90 (Hou et al., 2003), although Cs137 is easier to measure and — due to its long half-life — mainly responsible for the 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 3,474 measurement points in May 1986. The data were compiled by the BfS following a comprehensive measurement program rolled out between 1986 and 1989. Measurements taken after May 1986 were backdated based on the decay of CS137. For each municipality centroid, we calculate the radiation level as the inverse-distance weighted average from the four closest measuring points. After 1989, no comparable radiation data are available. Therefore, we know the *initial* level of radiation in area, but we have no information how radiation levels developed between 1989 and 2010. It is possible to calculate the approximate radiation level based on the decay of Cs137, although to determine the exact level we would need to know the extent to which the radioactive matter was washed into deeper layers of soil.

Linkage between individual and regional data We link the radiation data for 1986 with the individual survey data based on the respondents' municipality of residence in May 1986, using the radiation level in the centroid of the municipality. This linkage provides us with a measure of potential exposure to the post-Chernobyl radiation for each person in the sample.[9] Because we link the data without knowing the precise place of residence within a municipality, the linkage inevitably introduces measurement error. To address this problem, in Appendix B.2 we perform robustness checks, which show that the results are not driven by the linkage procedure.

Additional data We supplement our dataset with municipality- and county-level data on geographic conditions and population characteristics. We obtained data on precipitation, altitude and population size at the municipality level and data on minimum altitude and the composition of the population at the county level. In addition, we obtained data on dementia incidence from

[9] 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, the 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.

the German Hospital Quality Reports provided by Destatis. This dataset includes all dementia cases that were diagnosed in hospitals in a given municipality between 2006 and 2016. In Appendix A.4, we provide a detailed description of all variables used in this study.

4.5 Descriptive statistics

Table 1 displays the descriptive statistics of the main variables used in the regression. In 1986, the average person in the sample was 19 years old, with ages ranging from zero to 30 years. 36% of the sample — predominantly the older cohorts — were employed at the time, while another 43% were enrolled in education, and 1% were unemployed. The share of people who lived in the GDR represents 18% of the sample.

The German secondary school system has 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. 45% of the sample were no longer in education in April 1986: 4% had a lower secondary or secondary, while 28% and 13%, respectively, had an upper secondary or tertiary degree. On the other hand, 43% were still in education, most of whom had not yet finished a degree (31% of the sample). 10% of the sample were enrolled in 1986 but had already passed lower secondary or secondary education, while 1% had passed upper secondary 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 until the 1950s and earlier had much lower educational attainment than their children. Over half of all respondents have parents with no more than nine years of schooling.

The fourth set of statistics describe the cognitive test scores. Two features are noteworthy 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 and compare. Second, the number of observations differs between tests, which is due to design features of the NEPS (see Section 4.3 and Appendix A).

Panel B displays the municipality-level characteristics. With the exception of dementia incidence, the statistics were computed across individual observations in the estimation sample. The mean ground deposition of Cs137 in May 1986 amounts to 5.18 kBq/m^2. The standard deviation — which is larger than the mean — points to a significant variation in ground deposition across Germany.[10]

The level of precipitation represents the average rainfall in May in the five years preceding the Chernobyl disaster, i.e. 1981-1985. In 1986, the average person lived in a medium-sized municipality with 282,000 inhabitants, although municipality sizes vary between 5,000 and over

[10]See Appendix A.5 for an illustration of the distribution of the ground deposition across municipalities.

3 million. The last row of Panel B displays the number of dementia cases between 2006 and 2016 that were diagnosed in hospitals. Across municipalities, this number varies from zero to over 6,000, with a mean of 148.

Panel C lists county-level characteristics. Perhaps surprisingly, the share of people with a tertiary education in 1986 is very low and stands at 7%. The main reason behind this small number is that the German education system was traditionally based on vocational education, whereas university enrollment has only risen significantly since the 1980s. At the sample average, 44% in a county are working, whereas the share of working-age population is 65%.

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	4440
Female	0.51	0.50	0.00	1.00	4440
Native speaker	0.98	0.15	0.00	1.00	4440
GDR	0.18	0.39	0.00	1.00	4440
Unemployed in April 1986	0.01	0.12	0.00	1.00	4440
Employed in April 1986	0.36	0.48	0.00	1.00	4440
Educational attainment in April 1986					
Not of school age yet (less than 7 years old)	0.12	0.33	0.00	1.00	4440
No degree, lower secondary, secondary	0.04	0.19	0.00	1.00	4440
Upper secondary	0.28	0.45	0.00	1.00	4440
Tertiary	0.13	0.33	0.00	1.00	4440
In school or college education,	0.43	0.49	0.00	1.00	4440
no degree	0.33	0.47	0.00	1.00	4440
already attained lower secondary, secondary	0.10	0.31	0.00	1.00	4440
already attained upper secondary	0.01	0.09	0.00	1.00	4440
Highest parental education					
Lower secondary	0.52	0.50	0.00	1.00	4440
Secondary	0.27	0.44	0.00	1.00	4440
Upper secondary	0.21	0.41	0.00	1.00	4440
Test Scores					
Math	11.32	4.75	0.00	21.00	2652
Reading	27.06	7.45	0.00	39.00	2666
Reading Speed	38.19	8.34	0.00	51.00	3611
Scientific literacy	19.00	5.29	0.00	30.00	3286
ICT	41.20	13.62	0.00	66.00	3312
Reasoning	8.94	2.38	0.00	12.00	3169
Listening comprehansion	75.82	7.97	0.00	89.00	3172
Perceptual Speed	34.68	8.07	0.00	82.00	3170
B. Municipality-level data					
Caesium137 kBq/m^2 (01. May 1986)	5.18	5.87	0.50	62.10	4440
Average Caesium137 kBq/m^2 (until 2010, decay corrected)	3.89	4.41	0.38	46.64	4440
Precipitation mm/m^2 (yearly average, 1981-1985)	3.09	0.84	1.30	8.00	4440
Altitude in meter	201.59	176.69	0.00	850.00	4440
Minimum altitude in meter in county	138.73	139.78	-1.00	660.00	4440
Population/1000	281.67	676.43	5.00	3420.00	4440
B. County-level data					
Tertiary degree/Population	0.07	0.03	0.03	0.16	4440
Working population/Population	0.44	0.07	0.33	0.61	4440
18-65 years old/Population	0.65	0.03	0.59	0.71	4440

Notes: This table displays the descriptive statistics for the variables used in the analysis. 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 and in C at the county-level, although the statistics themselves are computed at the individual level. The statistics for dementia cases are measured at the municipality-level.

5 Empirical Strategy

In this section, we present the empirical model and the identification strategy. We pursue two complementary identification strategies, namely selection on observables and instrumental variables. Balancing tests suggest that the level of fallout in a person's municipality of residence is uncorrelated with a large number of observable characteristics. To alleviate concerns that our results are driven by unobserved heterogeneity, we apply an instrumental variable strategy, exploiting abnormal rainfall patterns within a critical window of ten days after the disaster. We perform several diagnostic tests to corroborate the exclusion restriction. Finally, we discuss challenges to statistical inference due to cross-sectional dependence and multiple hypothesis testing.

5.1 Empirical model

Our aim is to estimate the impact of potential exposure to the post-Chernobyl fallout on cognitive skills. For this purpose, we estimate the following empirical model,

$$y_{ims} = \alpha + \beta \; CS137^{86}_{ms} + \boldsymbol{X'_{im}}\gamma + \delta_s + \varepsilon_{ims}. \tag{3}$$

The cognitive test score y_{ims}, measured in and after 2010, of person i who resided in municipality m in state s in May 1986 is regressed on the ground deposition of Cs137 in the same municipality in May 1986. The vector $\boldsymbol{X_{im}}$ controls for pre-treatment characteristics of individuals and municipalities as well as design features of the survey. At the individual level, it includes controls for gender, a quadratic in age, a dummy for whether a person is a German native speaker, an indicator whether the person was born in Germany, parental education, education in 1986, and employment status in 1986. It also includes municipality and county characteristics, namely the average daily rainfall between 1981 and 1985, altitude, log population, as well as the share of the population in a county with a tertiary education, the share of people working and the share of people aged 18-65.[11] To capture features of the survey design, we control for the year in which a test was taken, as well as membership in one of four test groups.[12] In some specifications, we also control for state fixed effects, δ_s.[13] 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, the coefficient β measures the *total* effect of the radiation level in 1986 on cognitive test scores. A higher initial radiation in a municipality

[11] The controls for altitude include two variables, namely the altitude at the municipality centroid as well as the minimum altitude in a given county. The combination of these two variables has been shown to be a determinant of orographic rainfall (Houze, 2012), which in turn has been shown to increase the level of fallout (Yasunari et al., 2011). Appendix A.4 provides further details on the control variables.

[12] See Appendix A.2 for a description of the survey design.

[13] In line with the state borders of 1986, we treat the GDR as one state, which results in a total of twelve states. East Berlin is counted as part of the GDR, while West Berlin is considered a state of its own. The results are robust to fixed effects with all sixteen post-1990 states. These results are available on request.

in 1986 leads to a higher average radiation between 1986 and 2010, which in turn leads to a higher potential exposure to radiation. In addition, β contains the direct biological channels as well as behavioral responses to the fallout, such as changes in diet, exercise habits or internal migration. While we do not observe whether a higher potential exposure leads to a higher actual exposure, an estimate different from zero $\hat{\beta} \neq 0$ provides indirect evidence that it does.

Given the nature of our data and estimation strategy, we need to address several challenges to statistical inference. To account for potential cross-sectional dependence of the error terms, we cluster the standard errors at the county level. We also test for spatial autocorrelation in the cognitive skills index. Based on Moran's I, we fail to reject the null hypothesis of zero spatial autocorrelation for both variables.[14] In Appendix E, we undertake several steps to assess the robustness of our inference. We perform permutation tests, and allow for clustering at the state level by performing a bootstrap-t-procedure (Cameron et al., 2008). We also account for potential multiple hypothesis testing with a summary index test (O'Brien, 1984; Anderson, 2008) and a step-down adjustment of standard errors (Benjamini and Hochberg, 1995).

5.2 Identification

In an ideal experiment, we would randomly assign fallout levels across people, such that the level of fallout would be unrelated to determinants of cognitive skills ($E(\varepsilon_{im} \times CS137_{im}) = 0$) and β would be causally identified. However, in reality this identifying assumption may not be valid due to residential sorting. Although the German population could not anticipate the nuclear disaster and the local levels of fallout, it is possible that the level of fallout is correlated with the determinants of residential sorting. For example, urban areas receive less rainfall, while they also attract highly-skilled people, which may lead to a spurious correlation between skill levels and fallout.

This identification problem is prevalent in studies concerning the impact of pollution. Most studies on *in-utero* exposure overcome this problem by exploiting critical periods during which pollution affects a fetus. These studies typically compare people exposed to the pollutant during a critical period to people in the same locality who were also exposed, although not during the critical period. The confounding effect of residential sorting is differenced out with location fixed effects. However, for the effect that we want to uncover — the effect of constantly higher exposure over 25 years — this identification strategy is not suitable because there are not many pronounced critical periods during adolescence and adulthood. Instead, our identification needs to rely on cross-sectional variation. The estimating Equation (3) compares the cognitive skills of people with similar observables who were differentially exposed because they lived in different places at the time of the disaster.

To estimate a causal effect, we pursue two identification strategies. First, we control for

[14] After controlling for individual, municipality and county characteristics, Moran's I of the residuals with threshold distance 100km is $I = 0.001$, with a p-value of $p = 0.671$.

many potential determinants of residential sorting and show that conditional on these controls the sample is balanced on observable characteristics. Second, we apply an instrumental variable approach that exploits idiosyncratic rainfall in a critical time window after the disaster.

5.2.1 Selection on Observables

To corroborate the identifying assumption, in Table 2 we perform balancing tests whereby we compare pre-determined characteristics of people who lived in areas with above- and below-median levels of fallout in May 1986. To be conservative, we report conventional standard errors, which make discovering a significant difference more likely. The comparison of the raw data in Columns (1)-(3) indeed points to residential sorting. People with a low education as well as those with less-educated parents 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, are 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. As shown in Columns (4)-(6), 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 identifying variation to within the state of residence in May 1986. Conditional on these controls, the sample is balanced on all observable characteristics, which is why our preferred specification will include controls for municipality-level characteristics as well as state fixed effects. The state fixed effects also address an important institutional feature in the German education system, namely the fact that states have a high degree of autonomy in their education policy. This manifests itself in significant differences in PISA test scores between states. We account for these differences by including state fixed effects.

5.2.2 Instrumental Variable Strategy

While the sample is balanced on a large number of observables, our estimate of β may be confounded by unobserved differences between individuals. To further strengthen our causal identification, we apply an instrumental variable approach. Following a commonly-used method in physics (Pálsson et al., 2006), we predict the amount of fallout based on the interaction of two forces, namely the amount of rainfall while the plume was hanging above a place ($rain_m$) and the amount of radioactive matter in the plume in a given place ($matter_m$).[15] It is intuitive why both forces determine the level of fallout. The only way that an area can receive fallout

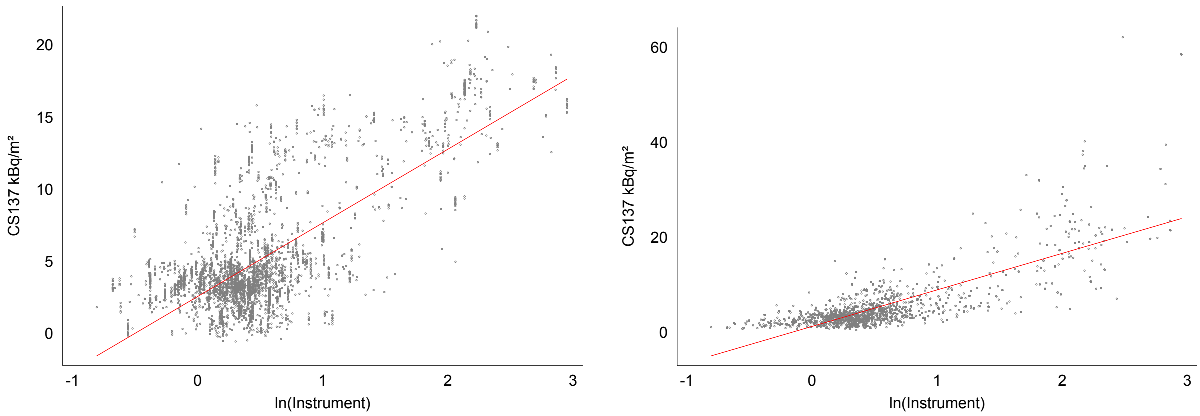
[15] We calculate the air concentration of radioactive matter based on measurements of sixteen measuring stations in Germany immediately after the disaster (April 29-May 8, 1986). Data on these measurements are provided by European Commission (1998). For each municipality, we compute the variable $matter_m$ as the inverse-distance-weighted average of the three closest measuring points.

is through rainfall: for a given amount of available radioactive matter in the plume, areas with higher rainfall receive more fallout. However, because the matter is rained down, the amount of radioactive matter changes as the plume moves. Given that large amounts of radioactive matter were rained down in the south of Germany, less matter could be rained down in North Germany.

The first stage In the first stage, we predict the level of CS137 based on the instrument — the log of the interaction of rainfall and radioactive matter as well as all the controls included in Equation (3),

$$CS137^{86}_{ms} = \lambda_0 + \lambda_1 \ln(rain_m \times matter_m) + \boldsymbol{X'_{im}}\boldsymbol{\kappa} + \rho_s + \eta_{ims}. \tag{4}$$

We measure $rain_m$ as the cumulative rainfall in a narrow time window between April 30 and May 8 in 1986 and interact this measure with the amount of radioactive matter in the plume. Importantly, in Equation (4) we control for the average level of rainfall between April 30 and May 8 in the years from 1981 to 1985.[16] Therefore, our instrument predicts the level of fallout based on *abnormal* rainfall levels, that is, the deviation in rainfall from the 5-year average within the critical ten days in early May.



(a) Raw correlation

(b) Correlation with controls

Figure 2: First-stage correlation

Notes: Panel (a) displays the first-stage correlation between the instrument $\ln(rain_m \times matter_m)$ and the amount of fallout CS137. In Panel (b) we control for the indivdual, municipality and county characteristics as well as state fixed effects mentioned in Section 5.1. The graph plots the residuals with the means of both axes added in.

Figure 2 displays the raw correlation between the instrument and the level of fallout (Panel a) and the correlation after controlling for indivdual, municipality and county characteristics and state fixed effects (Panel b). Even after adding controls, the correlation is strong and has the expected positive sign. In Table 13 in Appendix C.1, we report the first-stage coefficients

[16]Fine-grained rainfall data is only available from 1981 onwards.

and F-statistics for all outcomes and multiple specifications. In the first-stage regressions with all controls, the F-statistic of the excluded instrument ranges between 31 and 56, which rules out a weak instrument problem.

Exclusion restriction In order to deliver a causal estimate, the instrument has to satisfy the exclusion restriction that the rainfall on 10 days in early May 1986 — conditional on the *average* level of rainfall in early May and other controls — affects cognitive test scores only through its effect on fallout. It may well be possible that the average level of rainfall simultaneously affects the level of fallout and cognitive skills — for example if more intelligent people sort into places with less rainfall. However, this would not violate our exclusion restriction because we control for average rainfall. Our exclusion restriction would only be violated if deviation of rainfall in May 1986 from its average affected test scores through a channel other than fallout. We believe this is implausible given that we focus on a very short time window in early May. Abnormal rainfall in such a short period should not trigger behavioral changes that are so profound that they affect cognitive skills 25 years later unless this rainfall has a lasting effect on the environment.

To corroborate the exclusion restriction, we perform balancing and falsification tests. In the balancing tests, we regress pre-treatment characteristics on the instrument. A significant coefficient would indicate that the instrument is correlated with the error term in Equation (3) and, thus, invalid. The results, displayed in Table 14 in Appendix C.2, show that once we control for state fixed effects, the instrument does not predict pre-treatment characteristics. This finding is consistent with — although not a proof for — the instrument being as good as randomly assigned.

We further perform falsification tests based on the reduced form, i.e. a regression of the outcome on the instrument. Rather than using rainfall between April 30 and May 8, 1986, we compute the instrument based on rainfall on the same days in 1987 and 1988.[17] This is a diagnostic test for the validity of the instrument. If we found significant effects after 1986, this would indicate that our instrument picks up a rainfall pattern that is spuriously correlated with the outcome. This would be evidence against the exclusion restriction.

Table 15 in Appendix C.2 shows that while the reduced-form coefficients for 1986 are negative, large, and highly significant, they are considerably smaller, statistically insignificant and in most cases positive when the instrument is based on rainfall in later years. Among the 22 reduced-form regressions we perform, only one coefficient is statistically significant at the 5%-level, which is consistent with random sampling variation around a true effect of zero.

Further threats to identification. Besides residential sorting and unobserved heterogeneity, there are at least three additional challenges to identification. One challenge is selective attrition. Radiation can increase the risk of dying from cancer, potentially resulting in a se-

[17] We only use years after 1986 because we control for the average rainfall between 1981 and 1985 in all regressions.

lected estimation sample. Likewise, not all respondents completed all cognitive skills tests, and this non-participation is potentially systematic. Finally, 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 through as a series of robustness checks. We discuss the implications of these tests along with the main estimation results in the next section.

Table 2: Balancing tests based on pre-determined characteristics

	Raw data			Control municipality charac.			State FE, Municipality charac.		
	Mean		Diff (2)-(1)	Mean		Diff (5)-(4)	Mean		Diff (8)-(7)
	Below (1)	Above (2)	(3)	Below (4)	Above (5)	(6)	(7)	(8)	(9)
A. Individual characteristics									
Age in 1986	18.947	19.137	0.190	-0.290	0.273	0.563**	-0.189	0.177	0.366
	(0.178)	(0.170)	(0.246)	(0.178)	(0.168)	(0.245)	(0.178)	(0.168)	(0.245)
Female	0.517	0.499	-0.018	0.002	-0.002	-0.004	0.002	-0.002	-0.004
	(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	-0.002	0.001	0.003	-0.000	0.000	0.000
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Employed in April 1986	0.363	0.351	-0.012	-0.007	0.007	0.014	-0.003	0.003	0.006
	(0.010)	(0.010)	(0.014)	(0.010)	(0.010)	(0.014)	(0.010)	(0.010)	(0.014)
Unemployed in April 1986	0.014	0.013	-0.002	0.002	-0.002	-0.003	0.002	-0.002	-0.004
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
If employed: Qualified or highly qualified job before May 1986	0.519	0.524	0.005	-0.006	0.006	0.013	-0.005	0.004	0.009
	(0.017)	(0.017)	(0.025)	(0.017)	(0.017)	(0.024)	(0.017)	(0.017)	(0.024)
Children before 1986	0.178	0.154	-0.024**	-0.008	0.008	0.016	-0.005	0.005	0.011
	(0.008)	(0.008)	(0.011)	(0.008)	(0.007)	(0.011)	(0.008)	(0.007)	(0.011)
Older siblings	0.521	0.546	0.024	-0.000	0.000	0.000	-0.005	0.005	0.010
	(0.011)	(0.011)	(0.016)	(0.011)	(0.011)	(0.015)	(0.011)	(0.011)	(0.015)
Educational attainment in April 1986									
Lower secondary and secondary	0.032	0.043	0.011**	-0.002	0.002	0.003	-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	0.280	0.289	0.009	-0.004	0.004	0.007	0.001	-0.001	-0.002
	(0.010)	(0.009)	(0.014)	(0.010)	(0.009)	(0.014)	(0.010)	(0.009)	(0.013)
Tertiary	0.137	0.116	-0.021**	-0.002	0.002	0.004	-0.002	0.002	0.004
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
In school or college education	0.424	0.434	0.009	0.001	-0.001	-0.002	-0.002	0.002	0.003
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
In education, already attained lower secondary and secondary	0.103	0.104	0.001	-0.008	0.008	0.016*	-0.008	0.008	0.016*
	(0.007)	(0.006)	(0.009)	(0.006)	(0.006)	(0.009)	(0.006)	(0.006)	(0.009)
In education, already attained upper secondary	0.010	0.006	-0.005*	0.001	-0.001	-0.003	0.001	-0.001	-0.002
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
In education, already attained tertiary	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Not of school age yet (less than 7 years old)	0.127	0.119	-0.008	0.006	-0.006	-0.012	0.004	-0.004	-0.007
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
Highest parental education									
Lower secondary education	0.467	0.575	0.108***	-0.006	0.005	0.011	-0.006	0.005	0.011
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
Secondary education	0.288	0.249	-0.039***	-0.004	0.003	0.007	-0.002	0.002	0.004
	(0.010)	(0.009)	(0.013)	(0.010)	(0.009)	(0.013)	(0.010)	(0.009)	(0.013)
Upper secondary	0.246	0.176	-0.069***	0.009	-0.009	-0.018	0.008	-0.007	-0.015
	(0.009)	(0.008)	(0.012)	(0.009)	(0.008)	(0.012)	(0.009)	(0.008)	(0.012)
B. Municipality characteristics									
Caesium137 kBq/m^2 (01. May 1986)	2.283	7.905	5.622***						
	(0.014)	(0.150)	(0.150)						
Altitude in meter	141.419	258.079	116.660***						
	(2.852)	(4.048)	(4.952)						
Minimum altitude in meter in county	88.299	186.072	97.773***						
	(1.705)	(3.459)	(3.856)						
Population	438.286	134.627	-303.659***						
	(19.409)	(5.994)	(20.313)						
Precipitation in mm/m^2	2.911	3.264	0.353***						
	(0.021)	(0.014)	(0.025)						
GDR	0.289	0.082	-0.207***						
	(0.010)	(0.006)	(0.011)						
N	2150	2290							

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. Conventional standard errors of the test statistics are displayed in parentheses. Significance levels: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

6 Radiation and Cognitive Skills: Results

In this section, we present our estimation results for the effect of radiation on cognitive skills. We first show our baseline OLS estimates and discuss extensions such as non-linear effects and the effect of average exposure over 25 years. We briefly discuss the robustness with respect to estimation and inference, while we present a more detailed analysis in the appendix. To address concerns about endogeneity, we apply an instrumental variable strategy that exploits the fact that the radiation was driven by rainfall in a critical window of 10 days in May 1986. Finally, we compare our estimates to those from in-utero studies.

6.1 OLS Results

Table 3 displays the OLS results. Each coefficient is the result of a separate regression of the outcome listed on the left on the level of Cs137 in May 1986 and the controls listed at the bottom. The outcomes are standardized to mean zero and standard deviation one. A coefficient of $\widehat{\beta} = -0.01$ means that an increase in Cs137 by 1 kBq/m^2 is associated with a decrease in the respective test score by 1% of a standard deviation. To facilitate the interpretation, we discuss the effect sizes relative to an increase in CS137 by one standard deviation ($sd(CS137) = 5.87$).

Column (1) reports the OLS estimates from binary regressions without controls. All coefficients are small and statistically insignificant. In Column (2) we introduce controls for individual pre-determined characteristics. Although in some cases the sign switches, all coefficients remain close to zero. Matters are different in Column (3) when we control for municipality and county characteristics. After including these controls, the coefficients in Column (3) are considerably larger and in many cases statistically significant. A one-standard-deviation increase in the initial level of CS137 is associated with a decrease in the cognitive skills index by 4.1% of a standard deviation. The effect size ranges between tests from -0.6% of a standard deviation (logical reasoning) to -7.6% of a standard deviation (reading). In Column (4), when we include state fixed effects, the results are similar in magnitude and statistical significance.

The movement of the coefficients in Columns (1)-(3) of Table 3 may at first appear surprising. The balancing tests in Table 2 show that people with lower education as well as people with less-educated parents lived in areas that received more fallout. If education had a positive effect on cognitive test scores, one would expect a negative correlation between Cs137 and test scores. Once own and parental education is controlled for, one would expect the coefficient to become larger, i.e. less negative. However, in the German context it is plausible that we do not see the expected negative raw correlation. An important omitted variable here is the federal German education system, which varies in quality and institutions between states. The southern states of Bavaria and Baden-Württemberg traditionally had lower numbers of people with upper secondary or tertiary education, while at the same time their 9-th graders consistently achieved the highest standardized PISA test scores within Germany. Because both states have the highest average altitude and rainfall within Germany, they also received the largest amount of fallout.

The weak correlation in Column (1) can be the consequence of those two counteracting forces. The heterogeneity in education systems can also explain the coefficient movement between Column (2) and (3). If states with the highest altitude and rainfall levels offer the best education and if altitude and rainfall are positively related with the level of fallout, then the coefficient becomes more negative when we control for both variables.

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

	(1)	(2)	(3)	(4)
A. Individual test scores				
Math	0.003	-0.001	-0.011***	-0.011***
	(0.004)	(0.003)	(0.003)	(0.003)
Reading	-0.001	-0.005	-0.013***	-0.014**
	(0.006)	(0.005)	(0.005)	(0.005)
Listening comprehension	-0.003	-0.007**	-0.008**	-0.009**
	(0.004)	(0.003)	(0.004)	(0.004)
ICT	0.000	-0.002	-0.004	-0.005
	(0.002)	(0.002)	(0.003)	(0.004)
Scientific literacy	0.001	-0.000	-0.002	-0.003
	(0.003)	(0.002)	(0.003)	(0.003)
Reasoning	0.002	-0.002	-0.002	-0.001
	(0.003)	(0.003)	(0.004)	(0.004)
Reading speed	-0.001	-0.005*	-0.010***	-0.008**
	(0.003)	(0.003)	(0.004)	(0.004)
Perceptual speed	0.003	0.001	-0.003	-0.004
	(0.003)	(0.002)	(0.003)	(0.003)
B. Indices				
Cognitive skill index	0.001	-0.003	-0.008***	-0.008***
	(0.003)	(0.002)	(0.003)	(0.003)
Crystallized intelligence index	0.000	-0.003	-0.007**	-0.009**
	(0.003)	(0.003)	(0.003)	(0.003)
Fluid intelligence index	0.002	-0.002	-0.006**	-0.006*
	(0.003)	(0.002)	(0.003)	(0.003)
Controls:				
Individual characteristics	No	Yes	Yes	Yes
County characteristics	No	No	Yes	Yes
Municipality characteristics	No	No	Yes	Yes
State FE	No	No	No	Yes

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

Non-linear Effects In Appendix D.1, we test for a non-linear dose-response relationship. Based on a polynomial regression, a spline regression and a level-log regression, we find little evidence of non-linear effects.

The Effect of Average Exposure The estimates in Table 3 measure the effect of a higher initial potential exposure to radiation on cognitive skills. They are to be interpreted as reduced-form effects that summarize many channels through which cognitive skills are affected. However, the actual exposure over 25 years may be affected by at least two factors. First, due to the decay, the variation in radiation decreases over time. In addition, people who moved to an area with different radiation levels are exposed differently compared to people who stayed in their place of residence in 1986. In Appendix D we re-estimate the model in Column (4) of Table 3 but use as main regressors two measures of average exposure. The results are similar to the effects of the initial exposure. An increase in average exposure reduces cognitive test scores by between 3.2 and 9.6% of a standard deviation.

Robustness Checks In Appendices B and E, we assess the robustness of our estimation and inference. We show that the results are robust to different linking procedures of the radiation and survey data (B.2). We also rule out that the results are driven by selective mortality (B.3), design-based attrition or missing survey information (B.4). To account for multiple hypothesis testing, we perform a step-down correction of the p-values as well as a summary index test. Furthermore, we relax the parametric assumption of the normality of the error term and calculate the standard errors based on randomization inference. Finally, to allow for cross-sectional dependence within states, we perform a cluster bootstrap-t procedure (Cameron et al., 2008). Our inference is robust to these refinements of the calculation of the standard errors.

6.2 IV estimates

In Table 4, we present the instrumental variable estimates from regressions with all controls and state fixed effects. In Appendix C, we report the first-stage, reduced-form and second-stage results with varying sets of controls. Column (1) reproduces the OLS results from Column (4) in Table 3, whereas Columns (2) and (3) report the first-stage coefficients and F-statistics of the IV estimation.

The reduced-form estimates in Column (4) have the expected negative sign. More rainfall in early May 1986 means higher levels of fallout, which has a negative effect on cognitive skills. Five out of eleven coefficients are statistically significant at the 5%-level. While the magnitude of the reduced-form coefficient is not straightforward to interpret, the statistical significance is important for the interpretation of the second-stage results. It rules out that the second-stage results in Column (5) are a statistical artifact stemming from sampling variation in the first stage.

The second-stage results are larger than the OLS results. A one-standard-deviation increase in the fallout decreases cognitive skills by between 0% (perceptual speed) and 15% (reading) of a standard deviation. When we consider the overall cognitive index, an increase in fallout by one standard deviation leads to a decrease by 8% of a standard deviation.

There are three reasons why an IV estimator — provided that the instrument is valid and strong — can produce different results from an OLS estimator. First, the OLS estimates may be confounded due to unobserved heterogeneity that is not absorbed by our controls and fixed effects. Given the negative raw correlation between human capital and radiation, we would expect the OLS estimates to be smaller — that is, more negative — than the IV estimates. However, we observe the opposite result; our IV estimates are more negative than the OLS estimates. One possible explanation for this discrepancy is that the sorting on unobservable characteristics is positive, which would mean that people with higher unobserved levels of human capital lived in areas with higher radiation. However, to explain the large difference between OLS and IV estimates, selection on unobservables would have to be unrealistically large.

A second reason is measurement error in CS137, which can attenuate the OLS estimates. Given our data linkage procedure, measurement error is certainly present. We assign to each person the level of fallout at the centroid of his/her municipality of residence at the time of the disaster. However, the true exposure may differ from the assigned exposure because the level of radiation may vary within municipalities and because people differ in their lifestyles. The IV estimator absorbs the measurement error, which is one explanation why the estimates are larger.

Finally, the difference can be explained by heterogeneous treatment effects. If treatment effects are not constant across the population, the IV estimator identifies a local average treatment effect (LATE). With a continuous instrument, this means that estimator places a larger weight on municipalities with a higher level of compliance. In Appendix C.4, we explore potential sources of heterogeneous treatment effects. We find a stronger first-stage relationship in municipalities in southern states and in municipalities at above-median altitude. We also test for differences between urban and rural areas but find no significant difference. This suggests that the IV estimator places a larger weight on areas in the south and at higher altitude, which may deliver larger estimates than the OLS estimator, whose weights are determined by the state fixed effects.

6.3 Discussion of the main results and comparison with in-utero studies

The estimates presented in Tables 3 and 4 show that the radiation induced by Chernobyl had significant negative effects on cognitive performance. A one-standard-deviation increase in ground deposition reduces cognitive test scores between 4% and 8% of a standard deviation. With one percent of a standard deviation being roughly equivalent to the cognitive skills acquired in one school year, this means that receiving this additional radiation dose reduces a person's human capital by the equivalent of 4-8% of a school year.[18]

These effects appear economically significant when compared with the equivalent effective

[18] The equivalence between cognitive performance and school years is based on a regression of years of education on the cognitive skills index using the main estimation sample. We obtain a coefficient close to one.

Table 4: IV results: the effect of radiation on cognitive skills

	OLS	First stage	F-statistic	Reduced form	IV-2SLS
	(1)	(2)	(3)	(4)	(5)
A. Individual test scores					
Math	-0.011***	6.201***	66.162	-0.145**	-0.022**
	(0.003)	(0.041)		(0.058)	(0.010)
Reading	-0.014**	6.265***	46.565	-0.175***	-0.028***
	(0.005)	(0.086)		(0.040)	(0.008)
Listening comprehension	-0.009**	6.058***	58.930	-0.063	-0.010
	(0.004)	(0.048)		(0.048)	(0.008)
ICT	-0.005	6.619***	31.370	-0.029	-0.004
	(0.004)	(0.045)		(0.040)	(0.006)
Scientific literacy	-0.003	6.630***	31.219	-0.036	-0.005
	(0.003)	(0.047)		(0.040)	(0.006)
Reasoning	-0.000	6.056***	58.793	-0.039	-0.006
	(0.004)	(0.047)		(0.045)	(0.007)
Reading speed	-0.008**	6.403***	48.340	-0.105**	-0.016**
	(0.004)	(0.046)		(0.051)	(0.008)
Perceptual speed	-0.004	6.055***	66.162	-0.008	-0.001
	(0.003)	(0.046)		(0.046)	(0.007)
B. Indices					
Cognitive skill index	-0.008***	6.425***	42.264	-0.082**	-0.013**
	(0.003)	(0.035)		(0.037)	(0.006)
Crystallized intelligence index	-0.009**	6.420***	41.958	-0.094***	-0.015**
	(0.003)	(0.036)		(0.036)	(0.006)
Fluid intelligence index	-0.006*	6.220***	57.296	-0.048	-0.008
	(0.003)	(0.039)		(0.039)	(0.006)

Notes: This table displays the IV results. Column (1) reproduces the OLS results from Column (4) in Table 3. Columns (2) and reports the first-stage coefficients of Cs137 regressed on the instrument and all controls mentioned in Section 5.1 and the corresponding F-statistics, respectively. Column (4) reports the reduced-form coefficients of separate regressions of the outcomes on the instrument and controls. The main IV results are displayed in Column (5). Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

dose of other sources of radiation. Although the effective dose of Chernobyl is not straightforward to measure, estimates by the BfS suggest it is similar to the effective dose from medical procedures. The additional cumulative effective dose received by the average German over 25 years was around 0.6mSv, which is one-third of the effective yearly dose of background radiation (2mSv), or the equivalent of 30 chest x-rays. People in areas with higher contamination, for example Munich, received an effective dose of 2mSv, which is around the same as the dose from 150 chest x-rays or one CT scan of the head. Given that human cells react in a similar way regardless of whether a dose was received at once or over a longer period, our results suggest that low-dose radiation has an important effect on cognitive skills.

To further assess the magnitude of our results, it is useful to compare them with results obtained in studies exploiting in-utero exposure. The effect sizes vary across studies from large to very large. Heiervang et al. (2010) and Almond et al. (2009) estimate the effect of Chernobyl in Norway and Sweden, respectively. The average level of radiation in both countries was comparable to Germany; Norway had higher levels of variation in radiation across regions whereas the distributions in Sweden and Germany look similar. Heiervang et al. (2010) compare people exposed in-utero in the most- and least-affected areas of Norway and find a difference in IQ scores of 33% of a standard deviation. However, this is likely an overestimate, for people in the most-affected areas were recruited from rural areas whereas those in the least-affected areas came from the Oslo region. Part of the effect may be driven by unobserved differences between rural and urban areas.

The closest study for comparison is Almond et al. (2009). While their main specification is semi-parametric and, thus, difficult to compare, they also use the log level of fallout in some regressions. In Table 17, we estimate a similar specification and find that an increase in radiation by 100 log points reduced test scores by 7.6% of a standard deviation. The results in Almond et al. (2009) are significantly larger. They report an increase in math scores by almost 100% and an increase in overall GPA by 67.5% of a standard deviation. The effects found by Black et al. (2013) for Norway in the 1950s are considerably smaller. They report an effect between 2% and around 25% of a standard deviation.[19] One reason why these effects may differ is the fact that the contamination in Norway in the 1950s was only half of that in Sweden in the 1980s.

These comparisons suggest that the effects of post-natal exposure are an order of magnitude smaller than the effects of exposure during pregnancy. If cells are damaged while crucial body functions develop, it is unsurprising that they have larger effects than after birth, when this process has been finished. Nonetheless, our effects are economically significant, not least due to the relative number of people exposed after birth. In West Germany in the 1980s, the number of people between weeks 8 and 25 of gestation at any point in time was around 200,000. On

[19] The effect sizes in Almond et al. (2009) refer to the effects of $log(CS137)$ at the municipality-level reported in Table IV. The effect on math scores is -4.491, which is 96% of $sd(math) = 4.66$, reported in Table IX. The effect on GPA is -2.47, which is 67.5% of $sd(GPA) = 3.97$. Black et al. (2013) write on p. 24 of the NBER Working Paper version: 'Our log coefficients for IQ score are about -.04 for ground and about -.25 for air. These are approximately 2% and 12% of a standard deviation of the 25 dependent variables.'

the contrary, the size of the birth cohorts 1956-1985 was 24 million.[20] Therefore, the in-utero studies document a very large effect of an environmental shock on a small number of people, whereas our paper documents a smaller effect but for a population that is over 100 times larger.

7 Additional Results

In this section, we expand the analysis along several dimensions. We estimate heterogeneous effects, which reveal interesting differences between age groups and people of different socio-economic status. To provide further suggestive evidence on cognitive decline as a plausible biological channel, we estimate a fully flexible model of age-specific effects and document that radiation had an effect on the incidence of dementia more than 20 years later. Finally, we briefly describe evidence on behavioral responses (migration, education, labor supply), the details of which can be found in the appendix.

7.1 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. For example, in Column (1), we interact the ground deposition with a dummy for male and a dummy for female, which provides with separate estimates for both groups.[21] 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 the two 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. Upon first glance, this result seems surprising. The cohorts born in the first half of the 1980s were young children at the time of the disaster, and therefore were exposed in a critical phase of their development. In light of the literature on early-childhood exposure to pollution (Almond and Currie, 2011), we would expect the effect to be present among younger rather than older cohorts. Moreover, the works of Almond et al. (2009) and Black et al. (2013) show that children exposed to high radiation levels

[20] Source: vital statistics provided by Destatis.

[21] 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 both indicators in the regression. With only one indicator included — in this case a female dummy — the parameters are identified.

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. One potential explanation for this seemingly puzzling result is that the biological effects of *post-natal* exposure to radiation — those on brain cells and vital organs — are more likely to manifest themselves at older ages. However, because the youngest cohort was only 25 years old when they took the cognitive skills tests, we cannot observe what their test scores will be at age 50. Another potential explanation is that parents with young kids in 1986 particularly tried to shield their children away, thereby reducing the impact on later-life outcomes.

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 for people with less-educated parents is almost three times as large as the effect for those 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.

Finally, 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 unsurprising that the estimated effect in the GDR — although not statistically significant — is more than twice as large as the one for West Germany.

7.2 Evidence on Cognitive Decline

A plausible biological mechanism through which post-natal exposure to radiation affects test scores is cognitive decline. Constant exposure to higher radiation increases the likelihood that cells cannot reproduce. The cumulative effect over many years may result in lower cognitive test scores at older ages. At the same time, it is possible that high exposure during critical periods — for example puberty — has a stronger effect than during other periods. While our data and the cross-sectional setting do not allow us to explore these mechanisms in detail, we provide here two pieces of suggestive evidence.

Effects across age groups. First, we provide more evidence on heterogeneous effects across age groups. For this purpose, we estimate a difference-in-difference specification whereby we interact seven mutually exclusive dummies for people's age in May 1986 with the level of Cs137,

$$y_{itms} = \sum_{t=1}^{7} \gamma_t \times CS137_{ms} + \delta_t + \boldsymbol{X'_{im}} \boldsymbol{\gamma} + \delta_s + \varepsilon_{itms}. \tag{5}$$

Table 5: Heterogeneous effects

	(1)	(2)	(3)	(4)
CS137 kBq/m^2 $\times$ male	-0.008*			
	(0.003)			
CS137 kBq/m^2 $\times$ female	-0.008**			
	(0.003)			
CS137 kBq/m^2 $\times$ Age in 1986(0-10)		0.004		
		(0.005)		
CS137 kBq/m^2 $\times$ Age in 1986(10-20)		-0.018***		
		(0.007)		
CS137 kBq/m^2 $\times$ Age in 1986(>20)		-0.007**		
		(0.003)		
CS137 kBq/m^2 $\times$ Parent(above secondary education)			-0.004	
			(0.003)	
CS137 kBq/m^2 $\times$ Parent(below secondary education)			-0.010***	
			(0.003)	
CS137 kBq/m^2 $\times$ West Germany				-0.008***
				(0.003)
CS137 kBq/m^2 $\times$ East Germany				-0.017
				(0.015)
Controls:				
Individual characteristics	Yes	Yes	Yes	Yes
County characteristics	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Observations	4440	4440	4440	4440
R^2	0.22	0.23	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: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

The coefficients γ_t measure the effect of an increase in Cs137 by 1 kBq/m^2 on the cognitive skill index for the seven age groups $t = 1, \ldots, 7$, controlling for average differences across age groups (δ_t), state fixed effects (δ_t) and the same individual, municipality and county characteristics as before.

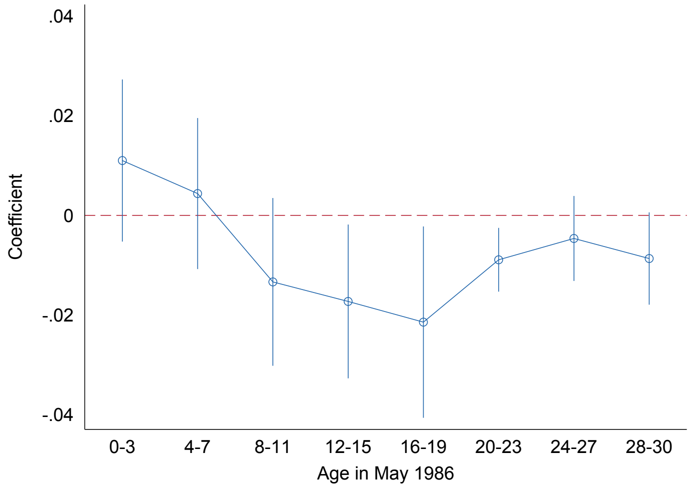


Figure 3: Differential effects by age group

Notes: This graph plots the estimated coefficients $\widehat{\gamma_t}$ and 95% confidence intervals for the age-specific effect of CS137 on the cognitive skills index. The estimates are based on Equation (5). Standard errors are clustered at the county level.

Figure 3 displays the estimated coefficients $\widehat{\gamma_t}$. While we find no significant effects for people aged 0-7 in May 1986, we find strong negative and in most cases statistically significant effects for ages 8 and older. The effects are strongest for people first exposed at ages 16-19. This can be seen as evidence that exposure matters most for cognitive skills during puberty, a time of significant re-modeling of the brain. However, we lack the statistical power to identify significant differences between age groups, such that the evidence is to be seen as suggestive.

Dementia diagnoses. As a second piece of evidence, we study the effect of radiation on dementia incidence. Recent evidence in the medical literature shows that low-dose radiation can increase the risk of dementia (Greene-Schloesser and Robbins, 2012; Kempf et al., 2016). Given that dementia is a cognitive dysfunction that manifests itself at older ages, it can be seen as an indicator for cognitive decline. To study the effect of radiation on dementia, we obtained data on dementia cases diagnosed in over 1,400 hospitals in Germany between 2006 and 2016. From the same data source — the German Hospital Quality Reports — we obtained data on diagnoses that should not be affected by radiation, namely asthma and injuries. These will serve as a placebo test.

Table 6 displays the estimated effects of an increase in radiation by 1 kBq/m^2 on the log number of diagnoses. The effect on dementia is positive once we control for geographic characteristics. In Column (3), when we condition on state fixed effects, the estimates become significant at the 5%-level. The point estimate of -0.077 means that an increase in radiation by one standard deviation increases the incidence of dementia by 45%. While this number sounds drastic, it has to be seen in relation to the mean of 148 diagnoses over an eleven-year period. An increase by 45% would mean an increase by 0.6 diagnoses per hospital per year. In Column (4) we use the same instrument as in the previous analysis. The first-stage relationship is different than in Table 4 because not all municipalities have a hospital. The IV estimate is larger than the OLS estimate. The point estimate of -0.12 is equivalent to a 70%-increase in the number of dementia cases for a one-standard-deviation increase in radiation. The results on asthma and injuries suggest that our estimates are unlikely to reflect a general health effect or some spurious relationship.

Table 6: Hospital diagnoses, 2006-2016

	(1)	(2)	(3)	(4)
Dementia	-0.017	0.041	0.077**	0.120***
	(0.025)	(0.031)	(0.033)	(0.046)
Asthma	0.025	0.050	0.055	0.042
	(0.062)	(0.064)	(0.061)	(0.062)
Injuries	0.048*	0.024	0.008	0.012
	(0.029)	(0.029)	(0.031)	(0.056)
First-stage:				
ln(Precipitation (mm/m^3) $\times$ Air contamination (kBq/m^3))				1.082***
				(0.113)
F statistic				91.169
Controls:				
County characteristics	No	Yes	Yes	Yes
Municipality characteristics	No	Yes	Yes	Yes
State FE	No	No	Yes	Yes

Notes: This table displays the estimation results for the effect of average ground deposition of Cs137 on the log cumulative number of hospital diagnoses between 2006-2016 at the municipality level. Each coefficient is the result of a separate regression of the variables on the left on the level of Cs137 in May 1986 and the controls listed at the bottom. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

7.3 Evidence on behavioral responses

In Appendix D.3, we test for behavioral responses along three margins, namely internal migration, employment and education. We run the same regressions as in (3), using as outcomes a binary indicator whether a person has moved until a given year, an indicator whether a person was employed in a given year and information on the number of years of completed education. We find evidence on neither of them. The effects on migration, employment and years of education are close to zero and precisely estimated. The only significant effect we find is on hours in continuing education — education people pursue while being employed. An increase in radiation by one standard deviation decreases the time spent in continuing education in 2010 by 9 hours (6.7% of the mean). Taken together, these results show little evidence that Chernobyl led to fundamental changes along these three margins.

8 Conclusion

In this paper, we show that radiation — even at subclinical doses — can have negative long-term effects on cognitive skills. Exploiting arguably exogenous variation in soil contamination in Germany after the Chernobyl disaster in 1986, we find that people exposed to higher radiation perform significantly worse in cognitive tests 25 years later. We find that the effect is stronger among older cohorts than younger cohorts, which is consistent with radiation accelerating cognitive decline as people get older.

These findings have implications for research and policy. Most research focuses on the effects of pollution exposure very early in life, often during pregnancy. Numerous studies show that exposure to pollution at this critical stage of a person's development has severe negative consequences. However, thus far there is little evidence of the impact of exposure *after* early childhood. By revisiting the consequences of the Chernobyl disaster with newly released data on adults' cognitive skills, we show that the negative effects of pollution are not limited to exposure early in life. Rather, we find the largest effects among people who were first exposed as adolescents. And while the effects of post-natal exposure are smaller than those of pre-natal exposure, they are economically significant nonetheless. This is particularly the case because the population exposed after birth was over 100 times larger than those exposed during the critical months of pregnancy.

For policy-makers, these results are important for at least two reasons. First, they point to substantial external costs of nuclear power generation. Although Chernobyl is over 1,000km away from the German border, the disaster's negative consequences significantly affect the German population. Indeed, while disasters like Chernobyl are rare, they certainly occur — for example, the Fukushima disaster in 2011 — and if they occur they come with serious negative consequences. Second, more generally, our results suggest that radiation has a human capital cost. While it is impossible for people to escape exposure altogether — natural radiation is present everywhere on earth — there are ways to shield the population away from it. One example is is through the choice of medical procedures. Analyses in the medical literature suggest that one-third of all CT scans are unnecessary (Brenner and Hall, 2007). Another example is the choice of building materials, given that some building materials are better at shielding people away from natural radiation, although their price may be higher than that of conventional materials. Our results can inform the cost-benefit trade-off of such choices.

This paper opens up several avenues for future research. Our results show that pollution can have negative long-term effects even if people are first exposed as adults. It will be important to understand if these results carry over to other pollutants such as particulate matter, ozone or lead. In addition, it will be important to obtain more accurate estimates of the magnitude of the impact of radiation. Due to data limitations, we are only able to measure a person's potential rather than actual exposure. While our setting allows us to obtain a causal estimate of an intent-to-treat effect, it would be useful by how much this effect would need to be scaled up to reflect the average treatment effect. Finally, the younger age cohorts in our sample seem too young for radiation to show its effect. As time goes by, it will be interesting to see if the effects of the younger cohorts are similar to those of the older cohorts.

References

Alinaghizadeh, H., Wålinder, R., Vingård, E. and Tondel, M. (2016). Total cancer incidence in relation to 137Cs fallout in the most contaminated counties in Sweden after the Chernobyl nuclear power plant accident: a register-based study. *BMJ Open* 6.

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) .

Arceo, E., Hanna, R. and Oliva, P. (2016). Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City. *The Economic Journal* 126: 257–280.

Aust, F., Gilberg, R., Hess, D., Kleudgen, M. and Steinwede, A. (2011). Methodenbericht: NEPS Etappe 8 Befragung von Erwachsenen Haupterhebung 1. Welle 2009/2010 .

Auvinen, A., Seppä, K., Pasanen, K., Kurttio, P., Patama, T., Pukkala, E., Heinävaara, S., Arvela, H., Verkasalo, P. and Hakulinen, T. (2014). Chernobyl fallout and cancer incidence in Finland 1988-2007. *International Journal of Cancer* 134: 2253–2263.

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). *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 Working Paper* .

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.

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.

Brenner, D. J. and Hall, E. J. (2007). Computed Tomography — An Increasing Source of Radiation Exposure. *New England Journal of Medicine* 357: 2277–2284.

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 . *NEPS Working Paper* .

Bundesamt für Strahlenschutz (2016). Der Reaktorunfall 1986 in Tschernobyl .

Bundesregierung (1986-1991). Bericht der Bundesregierung über Umweltradioaktivität und Strahlenbelastung .

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.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics & Statistics* 3: 414–427.

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

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

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M. and Rivkin, S. G. (2009). Does Pollution Increase School Absences? *The Review of Economics and Statistics* 91: 682–694.

Currie, J. and Neidell, M. (2005). Air Pollution and Infant Health: What Can We Learn from California's Recent Experience? *The Quarterly Journal of Economics* 120: 1003–1030.

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.

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.

Ebenstein, A., Lavy, V. and Roth, S. (2016). The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution. *American Economic Journal: Applied Economics* 8: 36–65.

European Commission (1998). Atlas of Caesium Deposition in Europe after the Chernobyl Accident .

Fielitz, U. and Richter, K. (2013). Bundesweiter Überblick über die Radiocäsiumkontamination von Wildschweinen .

Gehrer, K., Zimmermann, S., Artelt, C. and Weinert, S. (2012). The Assessment of Reading Competence (Including Sample Items For Grade 5 and 9) .

Gehrer, K., Zimmermann, S., Artelt, C. and Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online* 5: 50.

Gibbons, C. E., Serrato, J. C. S. and Urbancic, M. B. (2018). Broken or Fixed Effects? *Journal of Econometric Methods* 8.

Graff Zivin, J. and Neidell, M. (2012). The Impact of Pollution on Worker Productivity. *American Economic Review* 102: 3652–73.

Graff Zivin, J. and Neidell, M. (2013). Environment, Health, and Human Capital. *Journal of Economic Literature* 51: 689–730.

Greene-Schloesser, D. and Robbins, M. E. (2012). Radiation-induced Cognitive Impairment — from Bench to Bedside. *Neuro-Oncology* 14: iv37–iv44.

Haberkorn, K. and Pohl, S. (2013). Cognitive Basic Skills (Non Verbal) Data in the Scientific Use File .

Hachenberger, C., Trugenberger-Schnabel, A., Löbke-Reinl, A. and Peter, J. (2017). Umweltradioaktivität und Strahlenbelastung–Jahresbericht 2015 .

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.

Halla, M. and Zweimüller, M. (2014). Parental response to early human capital shocks: evidence from the Chernobyl accident. *IZA Discussion Paper* 7968.

Heiervang, K. S., Mednick, S., Sundet, K. and Rund, B. R. (2010). The Chernobyl Accident and Cognitive Functioning: A Study of Norwegian Adolescents Exposed In Utero. *Developmental Neuropsychology* 35: 643–655, pMID: 21038158.

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* .

Houze, R. A. (2012). Orographic effects on precipitating clouds. *Reviews of Geophysics* 50.

Ihme, J. M., Senkbeil, M. and Wittwer, J. (2015). The NEPS ICT Literacy Framework and Item Examples .

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. Tech. rep., International Atomic Energy Agency.

Kempf, S. J., Janik, D., Barjaktarovic, Z., Braga-Tanaka, I., Tanaka, S., Neff, F., Saran, A., Larsen, M. R. and Tapio, S. (2016). Chronic Low-dose-rate Ionising Radiation Affects the Hippocampal Phosphoproteome in the ApoE$-/-$ Alzheimer's mouse model. *Oncotarget* 7: 71817–71832.

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.

Lichter, A., Pestel, N. and Sommer, E. (2017). Productivity effects of air pollution: Evidence from professional soccer. *Labour Economics* 48: 54 – 66.

McCollough, C. H. and Schueler, B. A. (2000). Calculation of effective dose. *Medical Physics* 27: 828–837.

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.

National Council on Radiation Protection and Measurements (2009). Ionizing Radiation Exposure of the Population of the United States 160.

Nature (1992). Thyroid Cancer After Chernobyl. *Nature* 359: 21.

Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E. and Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for Educational Research Online* 5: 80.

O'Brien, P. C. (1984). Procedures for Comparing Samples with Multiple Endpoints. *Biometrics* 4: 1079–1087.

OECD (2016). Radiological Protection Science and Application .

Otake, M. and Schull, W. (1998). Review: Radiation-Related Brain Damage and Growth Retardation among the Prenatally Exposed Atomic Bomb Survivors. *International Journal of Radiation Biology* 74: 159–171.

Pálsson, S. E., Howard, B. J. and Wright, S. M. (2006). Prediction of spatial variation in global fallout of 137cs using precipitation. *Science of The Total Environment* 367: 745 – 756.

Pearce, M. S., Salotti, J. A., Little, M. P., McHugh, K., Lee, C., Kim, K. P., Howe, N. L., Ronckers, C. M., Rajaraman, P., Craft, A. W., Parker, L. and González, A. B. de (2012). Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet* 380: 499–505.

Peters, H. P., Albrecht, G., Hennen, L. and Stegelmann, U. (1987). Die Reaktionen der Bevï¿½lkerung auf die Ereignisse in Tschernobyl - Ergebnisse einer Befragung. Tech. Rep. Juel-Spez-0400, Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag, Jï¿½lich.

Preston, D. L., Shimizu, Y., Pierce, D. A., Suyama, A. and Mabuchi, K. (2003). Studies of Mortality of Atomic Bomb Survivors. Report 13: Solid Cancer and Noncancer Disease Mortality: 1950–1997. *Radiation Research* 160: 381–407.

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.

Renn, O. (1990). Public Responses to the Chernobyl Accident. *Journal of Environmental Psychology* 10: 151–167.

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.

Rumyantsev, P. O., Saenko, V. A., Ilyin, A. A., Stepanenko, V. F., Rumyantseva, U. V., Abrosimov, A. Y., Lushnikov, E. F., Rogounovitch, T. I., Shibata, Y., Mitsutake, N., Tsyb, A. F. and

Yamashita, S. (2011). Radiation Exposure Does Not Significantly Contribute to the Risk of Recurrence of Chernobyl Thyroid Cancer. *The Journal of Clinical Endocrinology & Metabolism* 96: 385–393.

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. Tech. rep.

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

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.

UNSCEAR (1994). Sources, effects and risks of ionizing radiation: UNSCEAR 1994 Report .

UNSCEAR (2000). Sources and effects of ionizing radiation: UNSCEAR 2000 Report .

UNSCEAR (2008). Sources and effects of ionizing radiation: UNSCEAR 2008 Report .

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

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. Tech. rep., Deutscher Wetterdienst.

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 .

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

A Data Description

A.1 Sampling in the ALWA subsample

As described in Section 4, 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, 250 municipalities were randomly sampled, and subsequently people were randomly sampled within municipalities. To make the sample representative, the number of people sampled within a municipality was proportional to the total population of the cohorts born between 1956 to 1986. Within municipalities, people's addresses were randomly sampled from person registers. This procedure resulted in a sample of 42,712 addresses, for which telephone numbers were collected. The telephone number of 22,656 people could be identified, and prospective participants were contacted by phone. Out of these, 10,404 actually completed the interview between August 2007 and April 2008, which corresponds to a response rate of 24.4% out of all sampled addresses, and 45.9% of all sampled telephone numbers.

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 their municipality of residence. If a person lived abroad, the name of the country of residence was collected. Municipality lists were provided to interviewers to ensure 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 highlighted spatial as well as chronological inconsistencies between the three domains (Drasch and Matthes, 2013).

Each life history module starts with a respondent's birth and further goes through their lives. In the case of 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

The NEPS is designed to assses competence development accross the lifespan starting with newborns (SC1), over pupils (SC2-SC4), students (SC5) to adults (SC6). All cohorts are tested along dimensions and tests are strongly oriented towards the concepts used by PISA. However, in order to make results comparable across cohorts, some adjustments were necessary, leading to deviations from the concepts used by PISA. Furthermore, the necessity of comparable test for children and adults explain the greater with PISA relative to other competence tests such as PIAAC. We explain the construction of all test dimensions covered in the SC6 in the following.

Reading competence The assessment of reading competence includes text functions like literary texts or advertising texts whereby 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 NEPS SC6. 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 PISA — with a focus on enabling participation in society.

However, the concept of reading competence in the NEPS distinguishes itself from PISA for two main reasons. In international studies of reading competence (e.g. PISA, CEFR), underlying texts are often categozized according to the type of situation in which they are applied — commonly with a focus on the reasons for reading such education, work, the personal domain and the public domain. However, reading competence in the NEPS is less oriented towards the reasons for reading, but rather it focuses predominantly on the functions of text along with the types of text associated with these functions, as well as how these relate to the cognitive requirements of reading. Furthermore, while PISA uses discontinuous texts, the 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 such as 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 the NEPS requires slightly different cognitive skills than the concept used in PISA, shown by tests measuring external validity (Gehrer et al., 2013).

Mathematical competence The test of mathematical competence comprises 21 items. Each item is equivalent to one point of the test score. The maximum processing duration is 28 minutes in a paper-pencil questionnaire (Schnittjer and Duchhardt, 2015).

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

As in the PISA mathematical competence test, the test in NEPS SC6 comprises 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 of external validity indicate a strong comparability with the same 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.

Based on 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 *scientific enquiry* and *scientific explanations* — which are required for personal decision-making. The maximum attainable test score is 28 points. The maximum processing duration is 25 minutes by paper-pencil questionnaire.

As in the PISA framework, the areas (KAS) and (KOS) are implemented in the context areas health, environment and technology. The concept of scientific literacy in the NEPS is slightly different from that of PISA due to time constraints in the number of items that can be asked within one test.

Listening comprehension This test analyzes receptive vocabulary. It measures the individual spectrum of vocabulary used in spoken language. Participants are provided with 89 items whereby they have to assign heard words to a sample of four pictures in front of them. The maximum attainable test score is 89 points. It follows the concept of the Peadbody Picture Vocabulary Test (PPVT) which is used in several large surveys such as the British Cohort Study, the European Child Care and Educaton Study (ECCE), or the National Longitudinal Study of Adolescent to Adult Health (AddHealth). For the SC6, the NEPS uses the publicly available German version of the PPVT published in 2004 (Berendes et al., 2013).

ICT Literacy *Information and Communication Technology (ITC) Literacy* includes components of computer literacy representing knowledge and skills necessary for the problem-oriented use of modern information technology.

This entails knowledge about basic operations, creating and editing documents as well as finding and assessing information. This test is in line with the literacy concept of PISA. The maximum test score is 68 points which can be attained in a maximum time of 25 minutes in a paper-pencil questionnaire (Ihme et al., 2015).

Reading speed The assessment of *reading speed* in the NEPS captures basic reading processes such as decoding, lexical access and basic sentence processing. The module comprises 51 short and simple statements. For each statement, the respondents have to assess if it is true or false. Therefore, the tests focuses on the automatized reading processes. The maximum attainable test score is 51. The test is based on the principles of the Salzburg reading screening (SLS) (Zimmermann et al., 2014).

Perceptual speed The test on *perceptual speed* reveals basic cognitive basic skills or, more precisely, the basal speed of information processing using picture symbol tests. The picture symbol test comprises two tables whereby in one of the tables each graphical symbol has a specific number. The second table displays the same symbols, although the corresponding numbers are missing. In the second table, participants have to find the numbers that equal the combination in the first table as fast as possible. within 90 seconds with a maximum of 93 items by paper-pencil questionnaires. This procedere follows the digit symbol coding of the Wechsler Adult Intelligence test (Brunner et al., 2014).

Reasoning Another test for cognitive basic skills is a matrix-based test which covers *reasoning*. It comprises nine items with several horizontally and vertically arranged boxes in which different geometrical symbols are shown. One field is left blank and has to be filled based on a logical series. The maximum attainable test score is 12 points. This procedere follows the matrix reasoning component of the Wechsler Adult Intelligence test (Brunner et al., 2014; Haberkorn and Pohl, 2013).

A.2.2 Cognitive indices

Along with separate regressions of each test score on radiation, we also analyze the impact of radiation on cognitive indices that summarize multiple dimensions of the latent factor cognitive skills. In order to compute each index, we first standardize each test score, then add the standardized test scores, and finally standardize this sum to mean zero and standard deviation one.

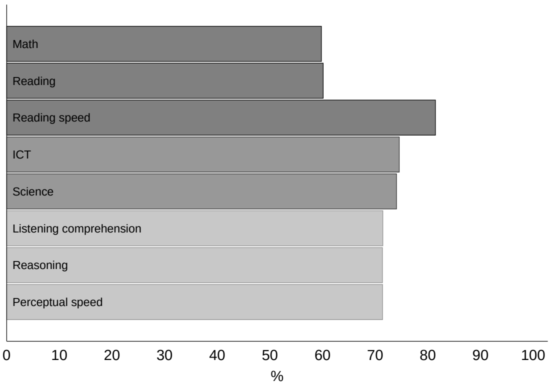
Besides an overall cognitive skill index that contains all eight test scores, we construct sub-indices for skills based on crystallized and fluid intelligence (Cattell, 1987). Research in radiobiology shows that radiation exposure has a larger effect on crystallized rather than fluid intelligence (Squire, 2009; Supekar et al., 2013). Following Salthouse (2012), we construct the fluid intelligence index based on the test scores of reading speed, perceptual speed and reasoning. The crystallized intelligence index comprises the five remaining test scores.

A.3 Participation in the competence tests

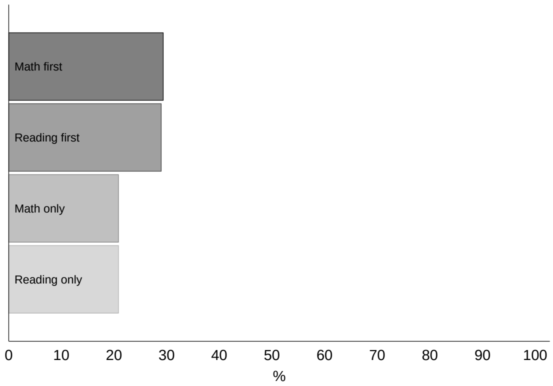
The tests were administered in three test periods between October 2010 and March 2015, namely tests in reading speed, math and reading between October 2010 and Mai 2011, tests in ICT and scientific literacy between October 2012 and April 2013, and tests in perceptual speed, listening comprehension and reasoning between August 2014 and March 2015. Most participants took their first test in the first test period, although, as illustrated in Figure 4c, some only started in the second and some few only in the third period.

As shown in figure 4b people were assigned to four different test groups which determined the test order. The test groups were created to decrease panel attrition by lowering participants' workload. In addition, the different test sequences ensure that the test results are not driven by the order in which the tests are administered. While the test order in the last period (2014/2015) was the same for all groups, it differed in the first two test periods in 2010/2011 and 2012/2013. Some test groups skipped one or more tests altogether. For example, reading was skilled in the third and math in the fourth test group.

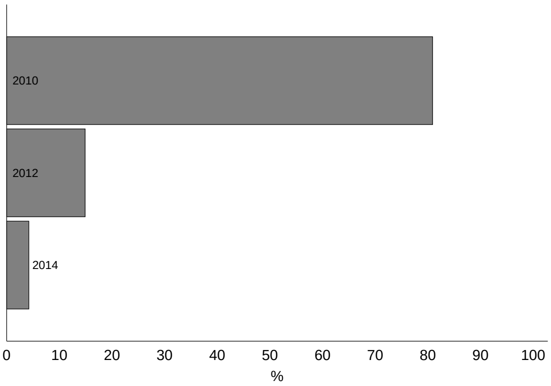
Figure 4a shows that participants do not necessarily perform all tests. The numbers of people completing a test varies between 2,644 (math test) and 3,602 (reading speed test). Overall, 4,423



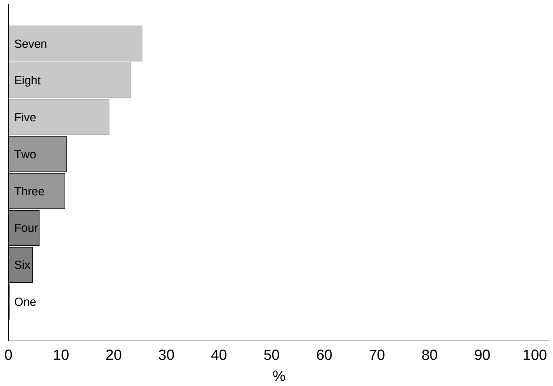
(a) Completed tests



(b) Test groups



(c) Year of first test



(d) Number of tests completed

Figure 4: Participation in cognitive tests

Notes: This figure displays descriptive statistics about the participation in cognitive tests for all 4,440 participants in our sample. Due to the survey design, not all participants took all tests, and tests were taken in different sequences. Panel (a) reports the share of participants who took a particular test. Panel (b) reports the distribution of test groups. Panels (c) and (d) show the distribution across years in which the first test was taken (left), as well as the number of tests taken by each participant (right).

participants performed at least one test. Figure 4d shows that most people completed at least seven tests, although a small number only performed one test. This difference in the number of tests completed is mainly due to the random assignment of people to tests. It is a design feature of the survey that not every participant had to complete all tests.

According to Aust et al. (2011), some participants refused to participate in competence tests. This was especially true for less educated participants. Furthermore, older people refused participation more often. In Appendix B.4, we test whether the non-participation in the competence tests is systematically linked to the level of radiation, which is not the case.

A.4 Regressor of interest and control variables

Table 7: Variables and Data Sources

Variable	Description
A – Individual-level Variables in NEPS	
Age in 1986	Continuous variable of participants' age in May 1986.
Female	Dummy variable of participants' gender: 1) Female 0) Male
Native speaker	Dummy variable of participants' first language: 1) German 0) Non-German
GDR	Dummy variable of participants' country of birth: 1) German Democratic Republic 0) Federal Republic of Germany
Unemployed in April 1986	Dummy variable of participants' unemployment status in April 1986: 1) Unemployed 0) Employed
Employed in April 1986	Dummy variable of participants' employment status in April 1986: 1) Employed 0) Unemployed
Not of school age yet (less than 7 years old)	Dummy variable of participants' enrollment status in April 1986: 1) Below 7 years old and not enrolled 0) 7 years old and above
No degree, lower secondary, secondary	Dummy variable of participants' educational achievements in April 1986 who are not enrolled but older than six years: 1) No degree, lower secondary, secondary 0) Others
Upper secondary	Dummy variable of participants' educational achievements in April 1986 who are not enrolled but older than six years: 1) Upper secondary 0) Others
Tertiary	Dummy variable of participants' educational achievements in April 1986 who are not enrolled but older than six years: 1) Tertiary 0) Others
In school or college education	Dummy variable of participants' educational activity in April 1986 who are older than six years: 1) Enrolled 0) Not enrolled

continued

Table 7 continued

Variable	Source
No degree	Dummy variable of participants' educational achievement in April 1986 who are older than six years and enrolled: 1) No degree 0) Others
Already attained lower secondary, secondary	Dummy variable of participants' educational achievement in April 1986 who are older than six years and enrolled: 1) Lower secondary, secondary degree 0) Others
Upper secondary	Dummy variable of participants' educational achievement in April 1986 who are older than six years and enrolled: 1) Upper secondary 0) Others
B – Municipality-level Variables	
Caesium137 kBq/m^2 (01. May 1986)	Continuous variable of the ground radiation of Caesium137 kBq/m^2 at the municipality of residence in 01. May 1986. We computed this variable for the municipality centroid based on the inverse-distance weighted average of the four closest measuring points. Source: Federal Office for Radiation Protection
Average Caesium137 kBq/m^2 (until 2010, decay corrected)	Continuous variable of decay corrected Caesium137 kBq/m^2 levels at the municipality of residence between 1986 and 2010. Decay formula: $Cs137_t = Cs137_0 \times e^{-0.024t}$, Source: Federal Office for Radiation Protection
Precipitation mm/m^2 (yearly average, 1981-1985)	Continuous variable of precipitation in mm/m^2, computed for the centroid of a municipality based on the inverse-distance weighted average of the four closest measuring points. Source: German Meteorological Service
Altitude in meter	Continuous variable of the municipality center's altitude. Source: Federal Agency for Cartography and Geodesy
Population/1000	Continuous variable of the municipalities' population in 1997 (in 1000). Source: Federal Agency for Cartography and Geodesy

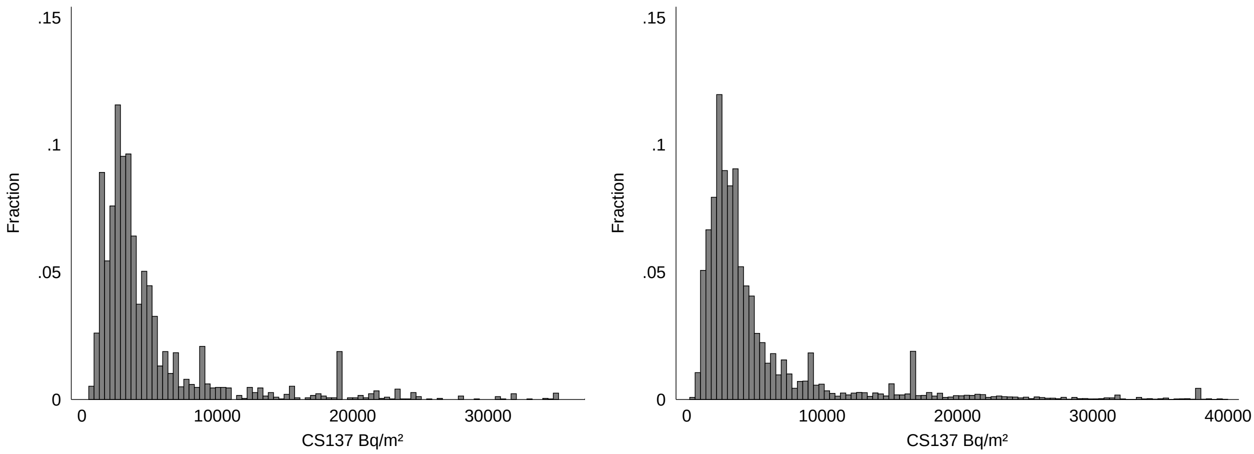
continued

Table 7 continued

Variable	Source
C – County-Level Variables	
Minimum altitude in meter in county	Continuous variable of the municipality centers' altitude that is the lowest in a county. Source: Federal Agency for Cartography and Geodesy
Tertiary degree/Population	Continuous variable of the population share in a county with tertiary degree. Source: Census Data of the GDR in 1987 and FRG in 1981.
Working population/Population	Continuous variable of the population share in a county that is working. Source: Census Data of the GDR in 1987 and FRG in 1981.
18-65 years old/Population	Continuous variable of the population share in a county aged between 18-65 . Source: Census Data of the GDR in 1987 and FRG in 1981.

A.5 Distribution of fallout

Figure 5 displays the distribution of the fallout in our sample as well as the German population. Based on participants' municipality of residence in May 1986, Panel (a) displays the the ground deposition of Cs137 in Bq/m^2 for the sample. Panel (b) shows the corresponding distribution for the entire German population, which we obtain by weighting the ground deposition in each municipality with the population in 1997. This was the first year for which consistent population data are available for the municipalities based on the same definition as the one used by the 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 municipality in 1997. Sources: Federal Office for Radiation Protection (Bundesamt für Strahlenschutz) and The Service Center of the Federal Government for Geo-Information and Geodesy.

B Robustness Checks

B.1 Goodness of fit for OLS regressions

Table 8 displays the adjusted R^2 for the OLS regressions in Table 3.

B.2 Robustness to different data linkage procedures

To generate our main regressor of interest, the amount of ground deposition in Cs137 in May 1986 in a person's municipality of residence at the time, it is necessary to link the radiation data with the survey data based on assumptions. While we have fine-grained data on Cs137 at a 3x3km grid-cell level, we only know a person's municipality of residence rather than the precise coordinates of their place of residence. In addition, the cell-level data have been generated by

Table 8: 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	0.19	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	0.19	0.20
Controls:				
Individual characteristics	No	Yes	Yes	Yes
County characteristics	No	No	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 (1)-(4) in Table 3.

the BfS based on an inverse-distance-weighted average of the four closest measuring points. In our main analysis, we link the radiation data via the geographic center (centroid) of each municipality. Both the interpolation by the BfS as well as the linkage via the centroid are potential sources of measurement error. Although we are unable to fully eliminate the measurement error, we can assess the robustness of our results to the choice of linkage procedure. In Table 9, we re-estimate the baseline model from Table 3, Column (8), with regressors based on different data linkages. The results in Table 9 strongly reject the notion that the results are driven our choice of linkage procedure.

- Column (1): baseline linkage, based on the inverse-distance-weighted average radiation of the four closest measuring points, linked via the municipality centroid
- Column (2): based on the radiation at the closest measuring point, linked via the municipality centroid
- Column (3): based on the inverse-distance-weighted average radiation of the four closest measuring points, linked via the population center of a municipality[22]
- Column (4): based on the radiation at the closest measuring point, linked via the population center of a municipality

[22] We computed the population center as the balancing point of a municipality based on night light data from 1996 provided by NASA.

- Column (5): based on the inverse-distance-weighted average radiation of the four closest measuring points, linked via the population mode of a municipality[23]
- Column (6): based on the radiation at the closest measuring point, linked via the population mode of a municipality
- Column (7): based on the unweighted average radiation in the entire municipality
- Column (8): based on the population-weighted average radiation in the entire municipality[24]

B.3 Testing for selective mortality

One important potential source of sample selection is selective mortality. Simply put, if radiation led to higher mortality among certain parts of the population, this population would be under-represented in our sample. To assess the importance of selective mortality, we obtained data on annual cohort-specific mortality data at the county level from the life tables of the German Statistics Office (Destatis).[25]. We run the following regression:

$$m_{crst} = \alpha + \rho_{rt}\ Cs137_{crs} + \boldsymbol{X'_{cs}}\boldsymbol{\kappa} + \delta_s + \varepsilon_{crst}. \tag{6}$$

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 δ_s. 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.

Figure 6a displays the estimates ρ_{rt} for all cohorts, while the remaining Figures present cohort-specific estimates. We find no evidence that exposure to the Chernobyl fallout led to higher mortality until 2010.

B.4 Testing for design-based attrition

As shown in the descriptive statistics in Table 1, not all respondents took part in all eight cognitive tests. This is mostly due to the random assignment of respondents into test groups, whereby some test groups skipped one or more tests. In addition, some respondents refused to take one or more tests. Such selection into competence tests could confound our results if systematically related to the ground deposition of Cs137. To test whether this is the case, we regress participation dummies (one if a person completed a test, zero if not) on Cs137 as well as the same controls as in our baseline regressions. As Table 10 shows, there is no evidence of systematic attrition or non-response once we add appropriate controls.

In Table 11, we provide additional evidence that observations with missing information are missing at random. In Panel A, the outcome is a dummy that equals unity if a person

[23] We take as population mode the point in a municipality with the highest light intensity in 1996.

[24] The averages in Columns (7) and (8) were computed based on the 3x3km grid-level data. construct the population weights, we used night light data from 1996.

[25] Such detailed data is only available from 1995 onwards

(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 standardized 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_{rt} for all cohorts in our estimation sample. Panels (b), (c), and (d) display the estimates of ρ_{rt} for distinct cohorts.

participated in at least one competence test. We regress this dummy on the level of Cs137 and in some specifications control for municipality characteristics and state fixed effects. The results strongly reject that non-participation in the competence tests is related to radiation exposure. In Panel B, we investigate whether non-response due to missing information is related to Cs137, but find no evidence. In Panel C, we test whether the random sampling of municipalities described in Appendix A was indeed random and therefore unrelated to the level of fallout. The results suggest that inclusion in the sample and the level of fallout are indeed unrelated.

B.5 The cognitive skills index with non-participation

Besides looking at the effect of radiation on separate cognitive tests, we also consider its effect on a cognitive skill index, which combines all eight test scores. To produce our baseline results, we computed the index regardless of the number of tests a person actually completed. This means that for some respondents the index is based on all eight test scores while for others it is based on just one. To assess whether the results are driven by non-participation, we re-estimate the baseline regressions but restrict the sample to all participants who completed at least a certain number of tests. Table 12 displays the results of this exercise. The coefficient in the first row is based on respondents who completed all eight tests, the coefficient in the second row is based on those who completed at least seven tests, the one in the third row is based on those who completed at least six tests, and so on. The coefficient in the last row represents our baseline estimate from Table 3, Column (8). The results show that, if anything, calculating the index based on all respondents leads to smaller estimates than calculating the index based on those who completed seven or eight tests.

Table 9: Robustness to the data linkage procedure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Individual test scores								
Math	-0.011***	-0.008***	-0.012***	-0.009***	-0.014***	-0.012***	-0.013***	-0.012***
	(0.003)	(0.002)	(0.005)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)
Reading	-0.013***	-0.008**	-0.017***	-0.010**	-0.017***	-0.007	-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*	-0.009**	-0.006*	-0.007	-0.005	-0.012***	-0.012***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.004)
ICT	-0.005	-0.002	-0.007*	-0.004	-0.003	0.001	-0.006*	-0.006*
	(0.004)	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Scientific literacy	-0.003	-0.002	-0.005	-0.003	-0.003	-0.002	-0.005	-0.004
	(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*	-0.010***	-0.004	-0.011**	-0.004	-0.010**	-0.009**
	(0.004)	(0.002)	(0.04)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Perceptual speed	-0.004	0.002	-0.005	-0.003	-0.004	-0.002	-0.005	-0.005
	(0.003)	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
B. Indices								
Cognitive skill index	-0.008***	-0.005**	-0.010***	-0.005**	-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	0.008**	-0.005**	-0.011***	-0.007**	-0.010**	-0.004	-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	-0.002	-0.007*	-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
County 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 estimation results whereby the regressor has been constructed with different data linkage procedures. See text in Section B.2 for a description of the linkage procedures. The controls are the same as in Table 3, Column (4). Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 10: Selection into competence tests

	(1)	(2)	(3)	(4)
Math	-0.002	-0.002	-0.001	0.000
	(0.002)	(0.002)	(0.002)	(0.000)
Reading	-0.000	-0.000	0.001	0.000
	(0.001)	(0.001)	(0.002)	(0.000)
Listening comprehension	0.000	0.002	0.003*	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
ICT	0.002**	0.003*	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	-0.002	-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:				
County characteristics	No	Yes	Yes	Yes
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: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 11: 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)	5844	5844	5844
B. Missing personal information			
Cs137 kBq/m^2	0.001	0.000	-0.000
	(0.001)	(0.001)	(0.001)
(N)	4545	4545	4545
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:			
County characteristics	No	Yes	Yes
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 that equals unity if a municipality was included in the NEPS SC6 sample and has at least one observation. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 12: The cognitive skills index with different definitions

	(Coef.)	(N)
All eight tests	-0.014**	1034
	(0.006)	
At least seven tests	-0.012***	2159
	(0.004)	
At least six tests	-0.013***	2360
	(0.004)	
At least five tests	-0.009**	3207
	(0.004)	
At least four tests	-0.010***	3466
	(0.003)	
At least three tests	-0.010***	3942
	(0.003)	
At least two tests	-0.008***	4430
	(0.003)	
At least one test	-0.008***	4440
	(0.003)	
Controls:		
Individual characteristics	Yes	
County 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: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

C Instrumental Variables: Details and Additional Results

C.1 First stages

Table 13 displays the first-stage coefficients for the regression Equation (4). Each coefficient is the result of a separate regression and is based on a different sample. The difference in samples is due to the fact that not every participant took every test, resulting in different first-stage coefficients and F-statistics. Column (1) displays the raw first-stage correlations without controls. In Column (2), we control for individual characteristics. In Columns (3) and (4), respectively, we add regional controls and state fixed effects. Introducing regional controls slightly reduces the size of the coefficients and the F-statistic decreases by about 50%. This indicates that part of the correlation of the instrument and the level of fallout is explained by regional factors such as population density, altitude and average rainfall. Nonetheless, even after introducing many controls, the F-statistic is above 30 in all regressions, which rules out a bias due to weak instruments.

C.2 Diagnostic tests in support of the exclusion restriction

In Table 14, we perform balancing tests by regressing individual pre-determined characteristics on the instrument and additional controls. Significant coefficients can be seen as evidence against the exclusion restriction as they suggest that the assignment of the instrument is not as good as random. In that case, it would be difficult to argue that the instrument is uncorrelated with the error term in Equation 3. The results in Table 14 suggest that the instrument passes this diagnostic test once we control for State fixed effects. The results support the identifying assumption that within states the assignment of the instrument is as good as random.

In Table 15, we perform an additional set of diagnostic tests by estimating the reduced form based on rainfall in different years. Ideally, we only want to find significant reduced-form effects based on rainfall in early May 1986 but not based on rainfall in early May in 1987 or 1988. Each coefficient is the result of a separate regression of the outcomes on the left on the instrument, individual-level controls, controls at the municipality- and county-level and state fixed effects. Column (1) displays the reduced form based on rainfall in 1986. All coefficients have the expected negative sign and 5 out of 11 coefficients are statistically significant at the 5%-level. In Columns (2) and (3) we estimate the same regressions but construct the instrument based on rainfall between May 1 and May 10, 1987 and 1988, respectively. Out of the 22 coefficients in both columns, one is significant at the 5%-level, which is consistent with random sampling variation. This indicates that the instrument works as it should. The assignment of the fallout is determined by rainfall while the plume was above Germany but not by rainfall on similar days in subsequent years.

C.3 Reduced-form and second-stage results

In Table 16, we report the reduced-form and second-stage results for different sets of controls. Each coefficient is the result of a separate regression of the outcomes on the left on the instrument (Columns (1)-(4)) or on the variation in CS137 that is predicted by the instrument (Columns (5)-(8)). Columns (1) and (5) display the results of univariate regressions. In Columns (2) and (6) we include individual-level characteristics. In Columns (3) and (7), we add municipality- and county-level controls, while in Columns (4) and (8) we add state fixed effects.

Table 13: First-stage coefficients

	(1)	(2)	(3)	(4)
A. Individual test scores				
Math	7.749***	7.667***	6.423***	6.483***
	(0.644)	(0.649)	(0.730)	(0.797)
F statistic	144.935	139.639	77.500	66.162
Reading	7.606***	7.505***	6.263***	6.265***
	(0.703)	(0.709)	(0.863)	(0.918)
F statistic	117.201	111.957	52.701	46.565
Listening comprehension	7.568***	7.474***	6.099***	6.058***
	(0.661)	(0.663)	(0.740)	(0.789)
F statistic	130.970	127.023	68.021	58.930
ICT	8.045***	7.944***	6.585***	6.619***
	(0.903)	(0.908)	(1.106)	(1.182)
F statistic	79.367	76.499	35.473	31.370
Scientific literacy	8.057***	7.952***	6.598***	6.630***
	(0.908)	(0.912)	(1.111)	(1.187)
F statistic	78.733	75.970	35.280	31.219
Reasoning	7.569***	7.474***	6.098***	6.056***
	(0.661)	(0.663)	(0.740)	(0.790)
F statistic	130.964	126.983	67.921	58.793
Reading speed	7.748***	7.663***	6.408***	6.403***
	(0.701)	(0.715)	(0.852)	(0.921)
F statistic	122.055	114.801	56.623	48.340
Perceptual speed	7.569***	7.474***	6.097***	6.055***
	(0.661)	(0.663)	(0.740)	(0.790)
F statistic	130.981	126.970	67.897	58.768
B. Indices				
Cognitive skill index	7.686***	7.594***	6.416***	6.425***
	(0.765)	(0.769)	(0.920)	(0.988)
F statistic	100.914	97.636	48.612	42.264
Crystallized intelligence index	7.684***	7.592***	6.406***	6.420***
	(0.767)	(0.771)	(0.923)	(0.991)
F statistic	100.242	97.013	48.178	41.958
Fluid intelligence index	7.548***	7.459***	6.120***	6.220***
	(0.653)	(0.660)	(0.764)	(0.822)
F statistic	133.709	127.842	66.307	57.296
Controls:				
Individual characteristics	No	Yes	Yes	Yes
County characteristics	No	No	Yes	Yes
Municipality characteristics	No	No	Yes	Yes
State FE	No	No	No	Yes

Notes: This table displays the first stage estimation results for the effect of ln(Precipitation (mm/m^3) $\times$ Air contamination (mm/m^3)) on Cs137. Each coefficient is the result of a separate regression of the variables on the left on a measure of ground deposition. Standard errors, clustered at the county level, are displayed in parentheses. The coefficients and F-Statistics differ because each estimation is based on a different sample owing to the fact that not every person took every test (see Appendix A). Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 14: Balancing tests for IV

	(1)	(2)	(3)	(4)
A. Individual characteristics				
Age in 1986	-0.220	-0.147	-0.302	-0.202
	(0.254)	(0.318)	(0.358)	(0.380)
Female	0.009	0.012	-0.015	-0.005
	(0.012)	(0.017)	(0.015)	(0.018)
Native speaker	0.009***	0.017***	0.014**	0.012
	(0.003)	(0.006)	(0.007)	(0.008)
Employed in April 1986	-0.001	0.016	0.003	0.002
	(0.014)	(0.019)	(0.017)	(0.021)
Unemployed in April 1986	0.001	-0.000	-0.004	-0.001
	(0.003)	(0.004)	(0.004)	(0.005)
If employed : Qualified or highly qualified	0.014	-0.016	-0.006	-0.006
	(0.023)	(0.032)	(0.034)	(0.035)
Children before 1986	-0.020**	0.003	-0.005	-0.005
	(0.010)	(0.013)	(0.011)	(0.014)
Older siblings	0.015	0.031	0.021	0.045
	(0.017)	(0.022)	(0.022)	(0.024)
Educational attainment in April 1986				
Lower secondary and secondary	0.004	-0.002	0.000	0.000
	(0.006)	(0.008)	(0.009)	(0.009)
Upper secondary	0.011	0.016	0.012	-0.000
	(0.011)	(0.017)	(0.015)	(0.019)
Tertiary	-0.009	0.003	-0.006	0.003
	(0.010)	(0.010)	(0.013)	(0.012)
In school or college education	-0.010	-0.026	-0.018	-0.013
	(0.014)	(0.019)	(0.018)	(0.021)
In education, already attained lower secondary and secondary	-0.002	0.001	-0.000	0.009
	(0.010)	(0.013)	(0.012)	(0.014)
In education, already attained upper secondary	-0.005***	-0.007*	-0.007*	-0.004
	(0.002)	(0.004)	(0.004)	(0.004)
In education, already attained tertiary	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Highest parental education				
Lower secondary education	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Secondary education	0.004	0.009	0.012	0.010
	(0.010)	(0.012)	(0.013)	(0.014)
Upper seconary	0.054***	0.011	0.024	0.024
	(0.016)	(0.020)	(0.021)	(0.023)
Controls:				
County characteristics	No	Yes	No	Yes
Municipality characteristics	No	Yes	No	Yes
State FE	No	No	Yes	Yes

Notes: This table displays the results of balancing tests for the instrumental variable. Each coefficient is the result of a separate regression of the variables listed on the left on the instrument $\ln(rain_m \times matter_m)$ and the controls listed at the bottom. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table 15: Diagnostic tests based on reduced form

	1986 (1)	1987 (2)	1988 (3)
A. Individual test scores			
Math	-0.145**	-0.042	0.014
	(0.058)	(0.089)	(0.089)
Reading	-0.175***	0.007	-0.060
	(0.040)	(0.077)	(0.092)
Listening comprehension	-0.063	0.082	0.021
	(0.048)	(0.070)	(0.086)
ICT	-0.029	0.052	0.060
	(0.040)	(0.065)	(0.073)
Scientific literacy	-0.036	0.128	-0.059
	(0.040)	(0.086)	(0.077)
Reasoning	-0.039	0.128*	0.166*
	(0.045)	(0.071)	(0.097)
Reading speed	-0.105**	-0.005	-0.023
	(0.051)	(0.076)	(0.085)
Perceptual speed	-0.008	-0.047	0.022
	(0.046)	(0.071)	(0.075)
B. Indices			
Cognitive skill index	-0.082**	0.046	0.010
	(0.037)	(0.063)	(0.066)
Crystallized intelligence index	-0.094***	0.056	-0.020
	(0.036)	(0.066)	(0.066)
Fluid intelligence index	-0.048	0.024	0.039
	(0.039)	(0.834)	(0.765)

Notes: This table displays the coefficients of reduced-form regressions of the outcomes listed on the left on the instrument. In each column we construct the instrument based on rainfall between May 1 and 10 in the year listed at the top of the table. In all regressions, we control for individual, municipality and county characteristics, as well as state fixed effects. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

The reduced-form coefficients in Columns (1) and (2) are small and statistically insignificant. They turn negative and are in most cases statistically significant once we control for regional characteristics (Column (3)) and add fixed effects (Column (4)). Given that the first stage is positive, the reduced-form results carry over to the second-stage coefficients in Columns (5)-(8).

C.4 Further evidence of a LATE

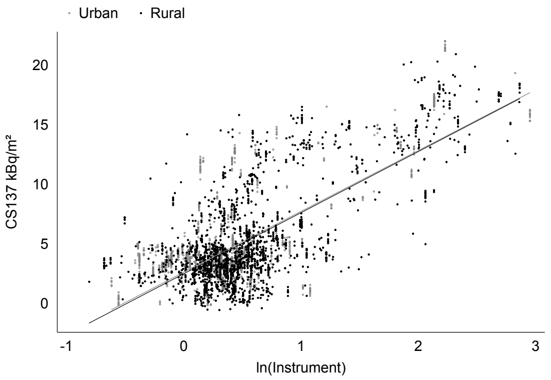
The analysis in Section 6 reveals robust negative effects, although the IV estimates are significantly larger than the OLS estimates. One reason for this difference is that both estimators apply different weights to observations and, thus, identify different effects. Because our model includes state fixed effects, the weights of the OLS estimator are determined by the number of observations per state and the within-state variance in treatment (Gibbons et al., 2018). The IV estimator, in contrast, identifies the local average treatment effect. With a continuous treatment, the estimator places higher weight on municipalities in which the instrument has a stronger effect on radiation.

In Figure 7, we provide evidence of differences in compliance between types of municipalities. Each panel plots the first-stage relationship with all controls and fixed effects for two distinct categories. In Panel a), we find similar first-stage relationships for areas with above- and below-median population density; both regression lines are virtually the same. In contrast, Panels b) and c) show significant differences. Panel b) plots the first stages for the southern states (Bavaria, Baden-Württemberg, Hesse and Rhineland-Palatinate) and the remaining states, labeled as 'north'. The first stage is strong in the south but not in the north. Similarly, Panel c) shows a stronger first-stage relationship for municipalities at above-median altitude. While not being an exhaustive list of categories, the three panels suggest it is plausible that the IV estimator identifies a LATE that gives larger weight to observations in southern states and those in municipalities at higher altitude.

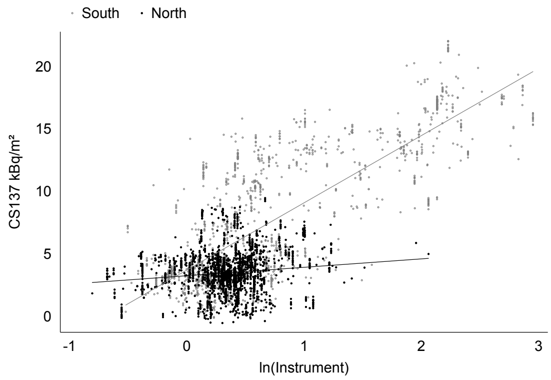
Table 16: IV second stage and reduced form

	Reduced forms				Second stages			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Individual test scores								
Math	-0.003	-0.025	-0.152***	-0.145**	-0.000	-0.003	-0.024**	-0.022**
	(0.040)	(0.036)	(0.052)	(0.058)	(0.005)	(0.005)	(0.009)	(0.010)
Reading	-0.035	-0.062	-0.178***	-0.175***	-0.005	-0.008	-0.028**	-0.028**
	(0.059)	(0.046)	(0.040)	(0.040)	(0.008)	(0.006)	(0.007)	(0.008)
Listening comprehension	-0.031	-0.044	-0.061	-0.063	-0.004	-0.006	-0.010	-0.010
	(0.045)	(0.041)	(0.045)	(0.048)	(0.006)	(0.006)	(0.007)	(0.008)
ICT	0.011	-0.026	-0.026	-0.029	0.001	-0.003	-0.004	-0.004
	(0.030)	(0.026)	(0.037)	(0.040)	(0.004)	(0.003)	(0.006)	(0.006)
Scientific literacy	-0.003	-0.013	-0.049	-0.036	-0.000	-0.002	-0.007	-0.005
	(0.032)	(0.028)	(0.040)	(0.040)	(0.004)	(0.003)	(0.006)	(0.006)
Reasoning	0.003	-0.041	-0.032	-0.039	0.000	-0.005	-0.005	-0.006
	(0.033)	(0.031)	(0.041)	(0.045)	(0.004)	(0.004)	(0.007)	(0.007)
Reading speed	-0.028	-0.069**	-0.167***	-0.105**	-0.004	-0.009**	-0.026**	-0.016**
	(0.037)	(0.031)	(0.051)	(0.051)	(0.005)	(0.004)	(0.008)	(0.008)
Perceptual speed	0.064**	0.016	-0.011	-0.008	0.008**	0.002	-0.002	-0.001
	(0.031)	(0.027)	(0.040)	(0.046)	(0.004)	(0.004)	(0.007)	(0.007)
B. Indices								
Cognitive skill index	-0.003	-0.034	-0.094***	-0.082**	-0.000	-0.004	-0.015**	-0.013**
	(0.034)	(0.028)	(0.034)	(0.037)	(0.004)	(0.004)	(0.006)	(0.006)
Crystallized intelligence index	-0.013	-0.036	-0.096***	-0.094***	-0.002	-0.005	-0.015**	-0.015**
	(0.035)	(0.030)	(0.033)	(0.036)	(0.005)	(0.004)	(0.006)	(0.006)
Fluid intelligence index	0.014	-0.025	-0.075**	-0.048	0.002	-0.003	-0.012**	-0.008
	(0.031)	(0.038)	(0.031)	(0.039)	(0.004)	(0.003)	(0.006)	(0.006)
Controls:								
Individual characteristics	No	Yes	Yes	Yes	No	Yes	Yes	Yes
County characteristics	No	No	Yes	Yes	No	No	Yes	Yes
Municipality characteristics	No	No	Yes	Yes	No	No	Yes	Yes
State FE	No	No	No	Yes	No	No	No	Yes

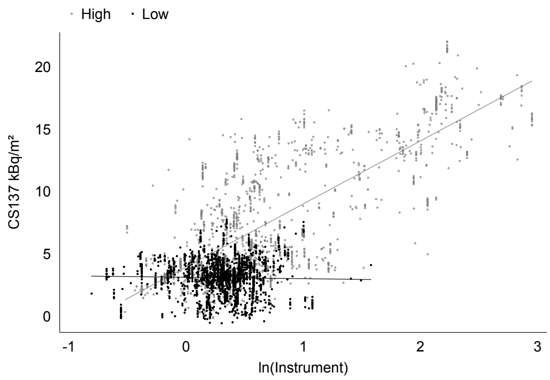
Notes: This table displays the reduced-form and second-stage estimations for the main specifications using ln(Precipitation (mm/m^3) $\times$ Air contamination (mm/m^3)) as instrument. In all regressions, we control for individual, municipality and county characteristics, as well as state fixed effects. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.



(a) Urban vs. rural areas



(b) South vs. north



(c) High vs. low altitude

Figure 7: First-stage correlations by category

Notes: This figure displays the scatter plots of first-stage regressions for different subgroups. In Panel (a), the sample is split between people living in municipalities with above- and below-median population density in May 1986; in Panel (b), the sample is split between states in the south (Bavaria, Baden-Württemberg, Hesse, Rhineland-Palatinate) and the remaining states (north); in Panel (c), the sample is split between municipalities at above- and below-median altitude.

D Additional Results

D.1 Non-linear effects

In Table 17, 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 (4) 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

Table 17: Non-linear effects

	(1)	(2)	(3)	(4)
CS137 kBq/m^2	-0.008***	-0.015**		-0.021
	(0.003)	(0.006)		(0.034)
CS137 kBq/m^2 $\times$ CS137 kBq/m^2		0.000		
		(0.000)		
ln(CS137 Bq/m^2)			-0.076***	
			(0.029)	
CS137 kBq/m^2 $\times$ above median				0.014
				(0.034)
Controls:				
Individual characteristics	Yes	Yes	Yes	Yes
County characteristics	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Observations	4440	4440	4440	4440
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 May 1986 in an area with an above-median ground deposition. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

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 is more appropriate 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 has a linear effect on test scores, which is why we use a linear model as our main specification.

D.2 The effect of average exposure

The coefficients in Table 3 measure the total effect of a higher ground deposition in 1986 on test scores 25 years later. The advantage of these estimates is that the initial level of ground deposition is arguably exogenous, such that these effects provide strong evidence of a negative causal effect of radiation on test scores. However, given that people were constantly exposed to the Chernobyl-induced radiation between 1986 and 2010, the interpretation of the magnitude is not straightforward. The most relevant measure for the regressor would be the cumulative effective radiation dose a person received during this period, namely the dose of radiation absorbed by all organs and tissues in a respondent's body. Obviously, such data would be very difficult to obtain as it would require measuring the energy absorbed by a person's tissue for every person in the sample.[26]

As a second-best solution, we consider as regressor the average ground deposition of Cs137 in the municipality where a person lived between 1986 and 2010, which measures a person's *potential* exposure to radiation in that period. To compute the average ground deposition in a municipality from 1986 to 2010, we calculate the ground deposition in every year based on the decay of Cs137 and take the average of all years. This measure serves as a proxy for a person's average potential exposure over 25 years.

In Table 18, we report the regression results for different measures of average exposure. All outcomes are standardized to mean zero and standard deviation one. As a benchmark, Column (1) reproduces the main regression results with the ground deposition in 1986 as the regressor. For years after 1986, we calculate the ground radiation based on the decay of CS137, $Cs137_{mt} = Cs137_{m0} \times e^{-0.024t}$ In Column (2), the regressor of interest is the average ground deposition of Cs137 in a respondent's municipality of residence in 1986. In Column (3), we take into account internal migration and use as regressor the average ground deposition in a respondent's municipalities of residence between 1986 and 2010. Due to the constant decay of Cs137, the variation in ground deposition across municipalities becomes smaller over time. As a result, the standard deviation of the average exposure is smaller than that of the initial exposure.

In Column (2), an increase in average ground deposition by 1kBq reduces the test scores by between 0.1% and 1.8% of a standard deviation. For the overall cognitive skill index, the effect is -1.1% of a standard deviation. Scaled up by the standard deviation of the average ground deposition ($sd = 4.41$), this is equivalent to a 4.9% of a standard deviation reduction

[26] Even in medical research, it is difficult to precisely measure the effective dose. Rather, the effective dose is estimated by simulation (McCollough and Schueler, 2000).

in the cognitive skill index for a one-standard-deviation increase in average ground deposition. The estimates in Column (3) are slightly smaller, ranging between 0 and -1.4% of a standard deviation reduction in test scores for an increase in ground deposition of 1kBq. Scaled up by the standard deviation of the average exposure ($sd = 3.23$), these effects range between 0 and -6.5% of a standard deviation, while the effect on the cognitive skills index is -3.2% of a standard deviation. In sum, these estimates have a similar magnitude to the reduced-form estimates in Column (1).

In Columns (4) and (5), we address three potential problems with our regressor in Column (3), the average ground deposition experienced by each person. The first problem is that people may move endogenously to avoid a higher radiation, which would bias the estimates. The second problem is measurement error in the regressor. We can only compute the amount of Cs137 in a given year based on its decay, but we do not observe the extent to which the radioactive matter is washed into deeper layers of soil. Because a person's exposure is higher the closer the matter is to the surface, our way of computing the average ground deposition inevitably introduces measurement error, which — if unsystematic — biases the results towards zero. A third problem is non-random sorting into areas, which may be correlated with the level of fallout. We address these problems by instrumenting the average ground deposition between 1986 and 2010 with the initial ground deposition in May 1986 (Column (4)) and with the instrument described in Section 5 (Column (5)). In both cases, the first-stage coefficient has the expected positive sign, suggesting that, on average, people who lived in areas with a higher ground deposition in 1986 were exposed to a higher average ground deposition over the following 25 years. The correlation is not perfect because people moved between places with different ground deposition. The instrumental variable estimates in both columns are considerably larger than the OLS estimates in Column (3). An increase in the average ground deposition by one standard deviation reduces cognitive test scores by 9.6% of a standard deviation. As described in Section 6.2, the IV estimates may be larger because of measurement error, unobserved heterogeneity and because the IV identifies a local average treatment effect. As shown in Appendix C.4, the IV estimator gives a higher weight to municipalities in the south than in the north and in general at higher altitude.

D.3 Evidence on behavioral responses

In Table 19, we explore the importance of several behavioral responses, namely internal migration, labor supply and investment in education. However, in our analysis we are constrained by the information available in our dataset. While the NEPS SC6 has rich information on some channels, we are unable to study several other behavioral responses such as changes in health behaviors, diet or exercise habits.

In the first panel of Table 19, we investigate 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 unsurprising, 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.

Table 18: The Effect of Average Exposure, 1986-2010

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)
Math	-0.011***	-0.015***	-0.014**	-0.026***	-0.049**
	(0.003)	(0.004)	(0.006)	(0.008)	(0.021)
Reading	-0.014**	-0.018**	-0.012*	-0.034***	-0.064***
	(0.005)	(0.007)	(0.007)	(0.012)	(0.016)
Listening comprehension	-0.009**	-0.012**	-0.012*	-0.023**	-0.026
	(0.004)	(0.005)	(0.006)	(0.009)	(0.018)
ICT	-0.005	-0.007	-0.003	-0.012	-0.011
	(0.004)	(0.005)	(0.005)	(0.008)	(0.014)
Scientific literacy	-0.003	-0.005	-0.001	-0.008	-0.014
	(0.003)	(0.004)	(0.005)	(0.008)	(0.014)
Reasoning	-0.001	-0.001	-0.000	-0.002	-0.016
	(0.004)	(0.006)	(0.007)	(0.011)	(0.017)
Reading speed	-0.008**	-0.010**	-0.019***	-0.018**	-0.037**
	(0.004)	(0.005)	(0.007)	(0.008)	(0.018)
Perceptual speed	-0.004	-0.005	-0.007	-0.010	-0.004
	(0.003)	(0.004)	(0.006)	(0.008)	(0.017)
B. Indices					
Cognitive skill index	-0.008***	-0.011***	-0.010*	-0.019***	-0.030**
	(0.003)	(0.004)	(0.005)	(0.007)	(0.013)
Crystalized intelligence index	-0.009**	-0.011**	-0.008*	-0.020**	-0.034**
	(0.003)	(0.005)	(0.005)	(0.008)	(0.014)
Fluid intelligence index	-0.006*	-0.008*	-0.011	-0.014*	-0.018
	(0.003)	(0.004)	(0.007)	(0.008)	(0.014)
First-stage: dep. var. Cs137 kBq/m^2					
Cs137 kBq/m^2				0.424***	
				(0.006)	
Precipitation (mm/m^3) $\times$ Air contamination (mm/m^3)					3.732***
					(0.304)
F statistic				912.733	53.902
Controls:					
Individual characteristics	Yes	Yes	Yes	Yes	Yes
County characteristics	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Mean (Cs137)	5.18	3.89	2.95	2.95	2.95
SD (Cs137)	5.87	4.41	3.23	3.23	3.23

Notes: This table displays the estimation results for the effect of average ground deposition of Cs137 from 1986-2010 on test scores. Each coefficient is the result of a separate regression of the variables on the left on a measure of ground deposition. Column (1) reproduces Column (4) in Table 3. In Column (2), the regressor is the decay-corrected average ground deposition from 1986 to 2010 in a respondent's municipality of residence in May 1986. In Columns (3) and (4), the regressor is the decay-corrected average ground deposition from 1986 to 2010, taking into account internal migration after 1986. In Column (4), we use the initial ground deposition in 1986 as an instrument for the average ground deposition between 1986 and 2010. In Column (5), we use the same instrument as in Section 6.2. The mean and standard deviation of Cs137 refer to the regressor used in each column. In all regressions, we control for individual, county and municipality characteristics, as well as state fixed effects. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

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. 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 reduces 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 of the behavioral responses that we are able to measure.

Table 19: Evidence on behavioral responses

	Coef.	(se)
Migration		
Until 1988	0.000	(0.001)
Until 1990	0.000	(0.002)
Until 1995	-0.003	(0.002)
Employment		
Month in employment between 1986 and 2010	3.027	(7.945)
Employed in 2000	0.000	(0.001)
Employed in 2005	-0.002	(0.002)
Employed in 2010	-0.001	(0.001)
Education		
Years in 1998	-0.004	(0.009)
Years in 1990	-0.008	(0.007)
Years in 1995	-0.009	(0.009)
Years in 2000	-0.008	(0.008)
Years in 2005	-0.005	(0.007)
Years in 2010	-0.007	(0.007)
Hours continuing education in 2010	-1.424	(0.574)**
Controls:		
Individual characteristics	Yes	
County 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: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

E Inference

E.1 Randomization inference

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

Figure 8a 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 only occurred once. The distribution in Figure 8b 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. This corresponds to an empirical p-value in a one-sided test of $p = 0.00026$.

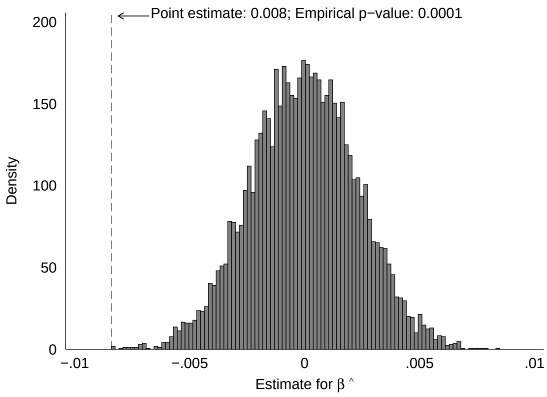
Figure 8c displays the placebo distribution of 10,000 estimates with randomization of the cognitive skill index across all observations. In 10,000 replications, such a result, equal to the point estimate in Table 3 in Column (3), only occurred 112 times, corresponding to an empirical p-value of $p = 0.0112$. The distribution in figure 8d corresponds to the estimations with state fixed effects presented in Table 3, Column (4). In this test, we randomize the outcome across observations within states and otherwise follow the same procedure as in Panel (c). A point estimate of -0.008 only occurs in 138 of 10,000 cases, corresponding to an empirical p-value of $p = 0.0138$.

In sum, these results reinforce the conclusions drawn from our inference with clustered standard errors in Section 6. If we consider the p-values of a two-sided hypothesis test — in which case the aforementioned p-values have to be multiplied by two — our estimates are statistically significant at the 5%-level.

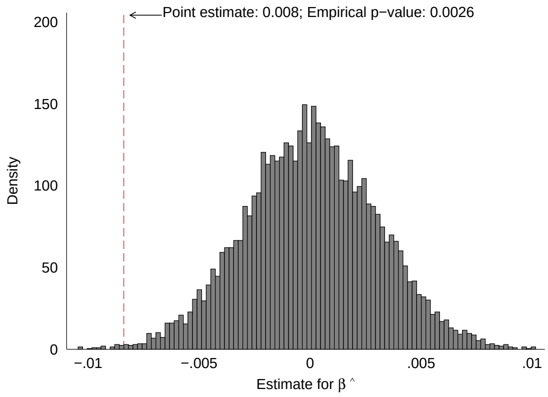
E.2 Multiple hypothesis testing

In our main analysis, we use eight cognitive test scores as outcome variables and estimate the impact of radiation on each outcome in separate regressions. However, because all of these outcomes represent different dimensions of the same latent factor cognitive skills, they are most likely correlated. This correlation leads to an underestimation of the standard errors and therefore an over-rejection of the null hypothesis. In other words, if the effect of radiation on one outcome is statistically significant, there is a high likelihood that the effects on other outcomes are statistically significant as well.

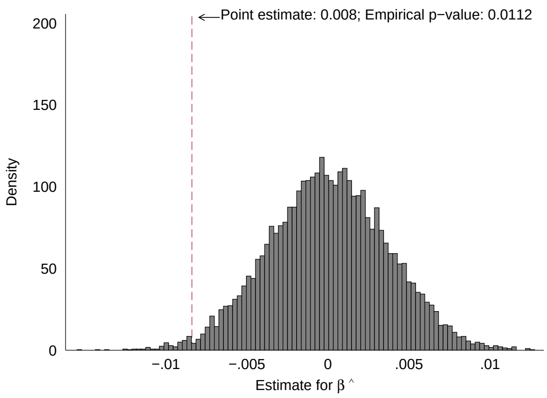
To take this correlation into account in the estimation of standard errors, the literature proposes two solutions. One is to keep the number of hypothesis tests constant but minimize the false discovery rate by adjusting the p-values. Another is to keep the p-values as they are but reduce the number of hypothesis tests, often to just a single test. In the following, we apply both approaches.



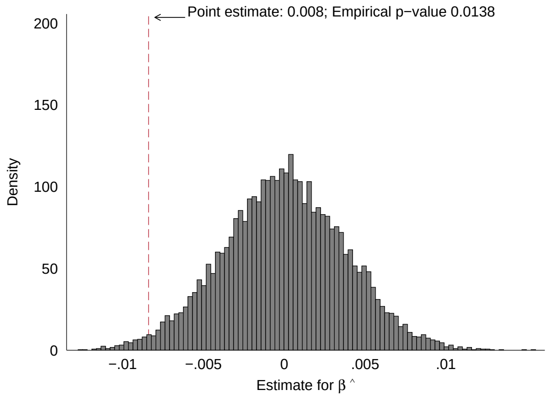
(a) Full randomization od CS137



(b) Randomization of CS137 within states



(c) Full rando. of cog. skill index



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

Figure 8: 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. In Panel (c), the outcome is randomized across all observations, whereas in Panel (d), the outcome is randomized across all observations within states. The vertical lines indicate the point estimate reported in Table 3 as well as the empirical p-values for one-sided tests.

To adjust the p-values, we follow the step-down approach by Benjamini and Hochberg (1995). This procedure is a refinement to the Bonferroni correction, in which p-values are adjusted by being multiplied with the number of hypothesis tests. The step-down approach assigns the largest adjustment to the p-value and the smallest adjustment to the highest. This approach is less conservative than the Bonferroni correction, which has been shown to cause severe under-rejection of the null hypothesis of no effect (Anderson, 2008).

Step-down approach In order to implement the step-down approach, we first rank all p-values from highest to lowest, and calculate the adjusted p-values — often referred to as q-values — using the formula

$$q = \frac{pm}{m - (i - 1)} \tag{7}$$

where p is the unadjusted p-value, m is the number of hypothesis tests, and i is the rank of the p-value, with $i = 1$ being the highest and $i = m$ the lowest. In our case, the highest p-value is unadjusted, whereas the lowest p-value is adjusted by a factor 8.

Table 20 displays the p-values and q-values for all eight outcomes.[27] After the adjustment, three coefficients remain statistically significant at the 5%-level and one (reading speed) at the 10%-level.

Table 20: Q-values (p-values adjusted by step-down approach)

	(1) p-values	(2) q-values
Math	0.001	0.006
Reading	0.010	0.041
Listening	0.021	0.057
ICT	0.156	0.250
Science	0.284	0.325
Reasoning	0.867	0.867
Reading Speed	0.028	0.057
Prectional Speed	0.210	0.280

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

Summary index tests A second approach that circumvents the problem of multiple hypothesis testing is to summarize all outcomes in a single index, in which case only a single hypothesis is to be tested and therefore no adjustment of the p-value is required (O'Brien, 1984; Anderson, 2008). The simplest index is constructed — as in our main analysis — by summing up the standardized outcomes and standardizing this sum. However, it is common practice to perform a summary index test on a weighted index, whereby each outcome is weighted by the additional variation that it contributes to the index. If a variable added to the index is highly correlated with a variable included in the index, this variable adds little new variation and thus receives a low weight. In practice, the weights are constructed from the inverted covariance matrix, whereby each outcome receives the sum of its row entries as a weight.

As shown in Table 21, the results only differ marginally between weighted and unweighted indices. Overall, these results — as well as those shown in Table 20 — confirm the statistical significance of the negative effect of radiation exposure on cognitive skills.

[27]Deviations of the q-values from Equation (7) are due to rounding.

Table 21: 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
County 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 (4), whereby the standardized indices are unweighted.

E.3 Cluster bootstrap-t procedure

In our baseline regression, we cluster the standard errors at the county level. However, this level of clustering may not be appropriate if the error terms are correlated between people living in different counties. For instance, this could be the case because in Germany education policy is set at the state level. However, adjusting for clustering at the state level with conventional cluster-robust standard errors can produce misleading results because the correction is based on the asymptotic assumption of the number of clusters going to infinity. With only sixteen states, this assumption is likely not fulfilled.

Cameron et al. (2008) provide a bootstrap-based method that allows for the calculation of standard errors with few clusters. Rather than sampling single observations in a bootstrap sample, this procedure samples entire clusters. Table 22 displays the main estimation results with standard errors, clustered at the state level, computed using the wild cluster bootstrap-t procedure. Compared to the conventionally-clustered standard errors — clustered at the county level — in Table 3, the bootstrapped standard errors are larger, although most estimates remain statistically significant at the 5%- or 10%-level.

Table 22: Estimates with cluster-bootstrapped standard errors

	(1)	(2)	(3)	(4)
A. Individual test scores				
Math	0.003	0.002	-0.011***	-0.011***
	(0.004)	(0.005)	(0.004)	(0.004)
Reading	-0.001	0.001	-0.013***	-0.013***
	(0.008)	(0.015)	(0.003)	(0.003)
Listening comprehension	-0.003	-0.003	-0.008**	-0.009***
	(0.003)	(0.003)	(0.004)	(0.003)
ICT	0.000	0.001	-0.003	-0.005*
	(0.003)	(0.006)	(0.003)	(0.003)
Scientific literacy	0.001	0.002	-0.002	-0.003
	(0.001)	(0.004)	(0.003)	(0.003)
Reasoning	0.002	0.001	-0.001	-0.001
	(0.004)	(0.004)	(0.005)	(0.006)
Reading speed	-0.001	0.001	-0.010*	-0.008*
	(0.007)	(0.079)	(0.005)	(0.004)
Perceptual speed	0.003	0.003	-0.003	-0.004
	(0.004)	(0.004)	(0.003)	(0.003)
B. Indices				
Cognitive skill index	0.001	0.002	-0.007**	-0.008**
	(0.008)	(0.006)	(0.004)	(0.004)
Crystallized intelligence index	0.000	0.002	-0.007**	-0.008***
	(0.004)	(0.005)	(0.003)	(0.003)
Fluid intelligence index	0.002	0.003	-0.006	-0.006
	(0.012)	(0.009)	(0.005)	(0.004)
Controls:				
Individual characteristics	No	Yes	Yes	Yes
County characteristics	No	No	Yes	Yes
Municipality characteristics	No	No	Yes	Yes
State FE	No	No	No	Yes

Notes: This table corresponds to the main regression table 3. The standard errors in this table have been computed based on the wild cluster bootstrap-t procedure by Cameron et al. (2008). Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

F Geographic information

In this section, we provide further background information on the measurement of radiation and rainfall, as well as the climatic conditions in Germany around the Chernobyl disaster.

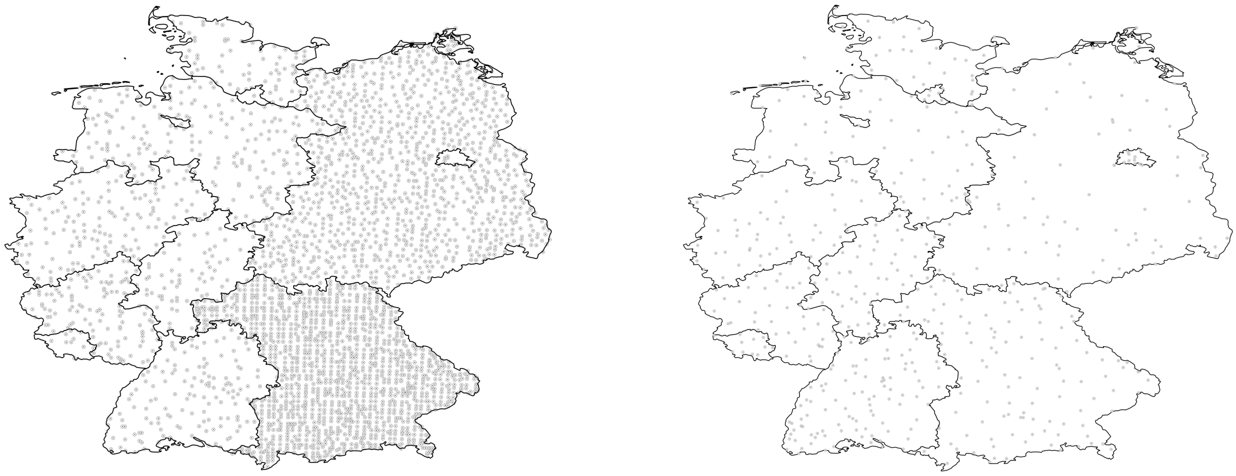
Measuring points for ground contamination Figure 9a shows the distribution of 3.448 measurement points for soil contamination, which is measured by an in-situ gamma ray spectrometer. Due the federal structure of Germany, several institutions were involved in the collection of measurements (Bavarian State 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.

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 9a 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).

Measuring points for rainfall Figure 9b shows the distribution of 544 weather stations. Coordinates as well as the rainfall data are provided by the German Meteorological Service. In the FRG, these stations are run by the German Meteorological Service. The stations in the GDR were operated by the Meteorological Service of the GDR which was eventually integrated in the German Meteorological Service. In comparison to Figure 9a, 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. For example, the inclination of the surrounding terrain should not exceed a specific limit, operation near high buildings is not possible and measurement operation should be executable for at least ten years (Wetterdienst, 2017).

Trajectory of the radioactive plume The radioactive plume reached Germany three days after the disaster, on April 30, 1986. It first entered the country in the south-east and made its way north-west before disappearing over the North Sea on May 10. The trajectory of the plume is illustrated in Figure 10, 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 south-east, close to the border with the Czech Republic and Austria, is located in the area



(a) CS137 (b) Precipitation

Figure 9: Measurement points. Source: The German Meteorological Service, The Federal Office for Radiation Protection

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.

Rainfall after the disaster The amount of precipitation Germany received between April 30 and May 8, 1986 is shown in Figure 11a. Darker color represents higher precipitation. We determine this period as critical period based on our observations in Figure 10. Comparing the level of precipitation with the ground deposition of Cs137 shown in Figure 1a, there appears to be a high correlation between the two. Figure 11b, in contrast, shows the average precipitation between 1981 and 1985. A comparison of Figures 11a and 11b, clearly shows that rainfalls in the critical nine days after the disaster introduced a high degree of idiosyncratic variation in rainfall and ground deposition. Some regions with traditionally high rainfall did not have any in those critical days, whereas some regions with traditionally low rainfall had exceptionally high amounts on these particular days.

Altitude and population density In the regressions, we control for altitude and population density, two potential determinants of both ground contamination and test scores. Figures 12a and 12b display the distribution of both variables across space.

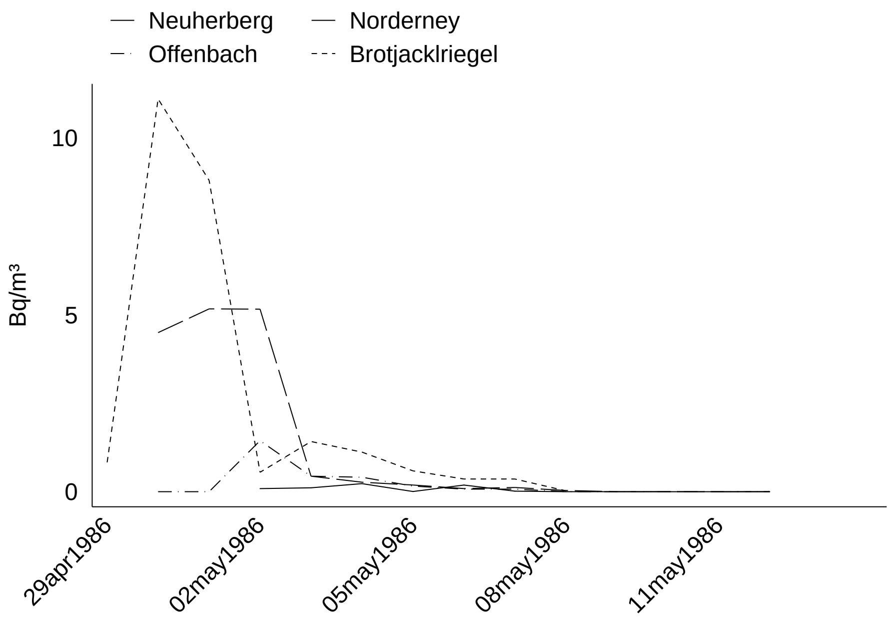
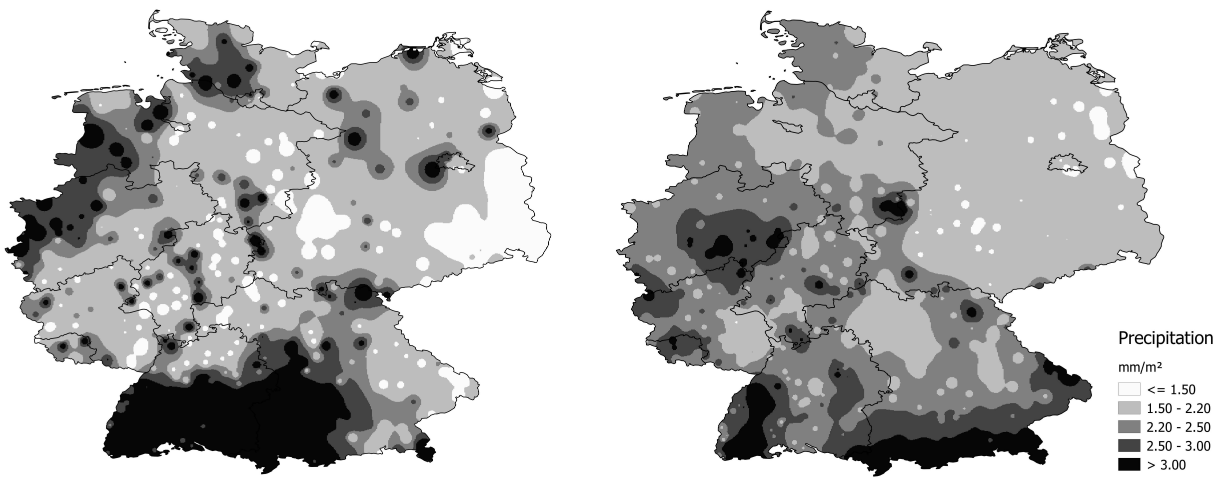


Figure 10: Air concentration of radioactive particles in 1986

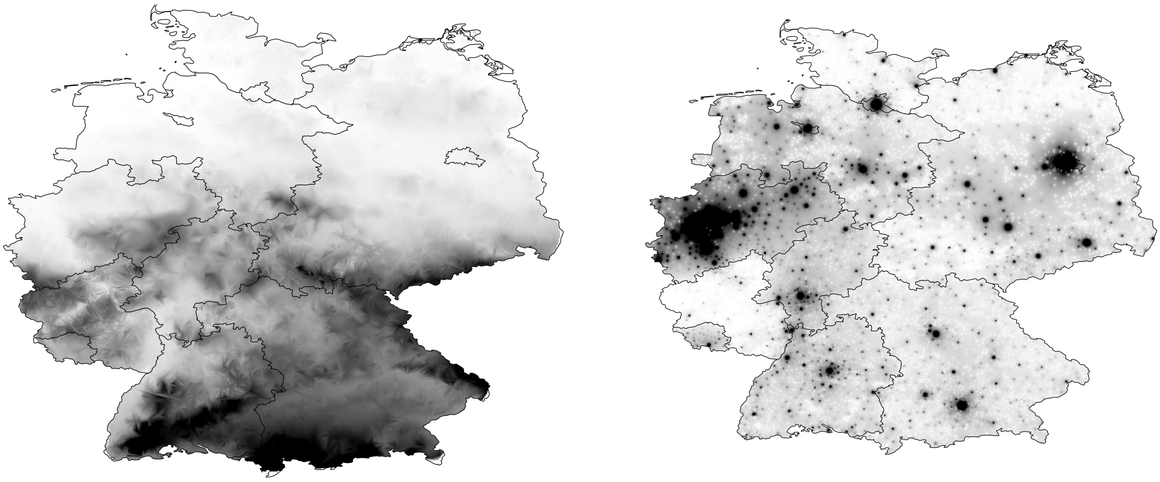
Notes: This graph displays the air concentration of Cs137 measured after the arrival of the radioactive plume in four German measuring stations. These are located in different parts of the country: Brotjacklriegel (south-eastern border), Neuherberg (south-east), Offenbach (center) and Norderney (north-west).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



(a) Critical 10 days in 1986

(b) Between 1981 and 1985

Figure 11: Average daily Precipitation. Source: The German Meteorological Service



(a) Topography

(b) Population density

Figure 12: Altitutde and population density, darker means higher. Source: Federal Agency for Cartography and Geodesy