

The Human Capital Cost of Radiation: Long-run Evidence from Exposure outside the Womb

Benjamin Elsner
University College Dublin, IZA, and CReAM

Florian Wozny
IZA

12 March 2018

Abstract:

This paper studies the long-term effect of radiation on cognitive skills. We use regional variation in nuclear fallout caused by the Chernobyl disaster in 1986, which led to a permanent increase in radiation levels in most of Europe. 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 (born before 1976), whereas we find no effect for people who were first exposed during early childhood. These results are consistent with radiation accelerating cognitive decline during older ages. Moreover, they suggest that radiation has negative 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, Joan Costa-i-Font, Thomas DeLeire, Olivier Deschênes, Paul Devereux, Jon Guryan, Adam Isen, Ingo Isphording, Shelly Lundberg, Keith Meyers and Nico Pestel, as well as audiences at IZA, U Freiburg, the 5th IZA Workshop on Environment and Labor Markets, RWI Essen, the 2nd IZA Workshop on the Economics of Education, U Regensburg, U Cologne and the UCD Geary Institute for helpful comments. 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 8505, 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 American receives almost twice the annual dose of radiation compared with 1980. This increase is almost entirely due to man-made sources of radiation such as CT scans, x-rays or mammograms, as well as flying, during which people are exposed to cosmic radiation (Brenner and Hall, 2007; Bolus, 2013). However, while there is no doubt about the health consequences of very high doses of radiation — those received by survivors of an atomic bomb or a nuclear disaster — the literature has reached no consensus on the consequences of small (subclinical) doses of radiation. Several studies find negative effects on health and cognitive performance when people are exposed *in-utero* (Almond et al., 2009; Heiervang et al., 2010; Black et al., 2013), although the evidence is less clear for people exposed as children or adults.¹ In this paper, we provide evidence of significant negative effects of subclinical radiation on cognitive skills. In particular, we show that these effects even exist when people are first exposed as adults.

To identify a causal effect, we exploit regional variation in the level of nuclear fallout of the Chernobyl disaster in 1986, after which large amounts of radioactive matter were spread across Europe. We focus on Germany, which — despite being over 1,000km from Chernobyl — received a large share of the fallout. Within Germany, the amount of fallout in a region crucially depended on rainfall within a ten-day window. In Munich, where rainfall was strong during that period, the ground contamination was seven times higher than in Hamburg, where it rained very little. Due to the long half-life of the radioactive matter, a high initial level of fallout led to a quasi-permanent increase in radiation levels in a region. People who lived in highly-contaminated areas in 1986 have been exposed to higher radiation until today. The radiation dose received during our sample period between 1986 and 2010 is comparable to that from other man-made sources. The average German received a dose of 0.6mSv — the equivalent of 30% of a CT scan, 1.5 mammograms, or 30 chest x-rays — although in more heavily-contaminated areas the dose was over 2mSv.

Our study is based on the National Educational Panel Study (NEPS), a representative survey of the German population born between 1956 and 1986. Three features of the NEPS are key to our analysis. First, for every respondent, the survey contains a detailed residential history, allowing us to link personal information with data on radiation in the respondent’s place of residence in 1986. Second, the survey includes eight standardized cognitive tests — math, reading, listening, reasoning, etc. — taken by the respondents when they were between 24 and 58 years old. Therefore, the NEPS is one of the few datasets with information on cognitive skills after school-leaving age. Finally, at the time of Chernobyl, around half of the sample were adults, which allows us to study the long-term effect of radiation when people are exposed as

¹Among radiotherapy patients, Hall et al. (2004) and Pearce et al. (2012) find negative effects of radiation on cognitive performance, whereas Blomstrand et al. (2014) find no significant effect. Alinaghizadeh et al. (2016) find a small effect of the Chernobyl fallout on the incidence of cancer in Sweden.

adults. We link 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 both datasets allows us to run a reduced-form regression of cognitive skills in 2010 and after on the initial level of fallout in a person’s place of residence in 1986.

Despite the arguably exogenous variation in fallout across regions, our identification faces several potential challenges, namely anticipation, residential sorting and endogenous migration. Our research design addresses all such issues. Anticipation — which is usually a challenge in studies on the impact of pollution — is an unlikely confounder because Germans could not anticipate the disaster, and most of the population only learned about it after the nuclear rain had fallen. Moreover, the data show that once we condition on state fixed effects, there is no apparent pattern of residential sorting. We also find no correlation between the initial fallout and the likelihood of subsequent internal migration.

Our central finding is that people exposed to higher radiation have significantly lower cognitive test scores 25 years later. A one-standard-deviation higher initial exposure in 1986 reduces test scores by 4.5 percent of a standard deviation. This means — for example — that a person exposed to the initial radiation level in Hannover would have 6.6% of a standard deviation lower test scores in 2010 compared to a person exposed to the level in Hamburg, all else being equal. Because our reduced-form regression identifies a combination of biological effects and compensating behaviors, this estimate is likely a lower bound to the pure biological effect of radiation on cognition. We also estimate the effect of the average exposure between 1986 and 2010, finding similar effects. An increase in average exposure by one standard deviation reduces cognitive test scores between 3 and 6%.

We further investigate whether the effect differs across cognitive domains and between demographic groups. We find that radiation has a stronger effect on skills relying on crystallized intelligence, such as scientific knowledge, math, reading or ICT skills, and a weaker effect on those relying on fluid intelligence such as reading speed, reasoning and perceptual speed. These results are consistent with the medical literature, which shows that radiation mainly affects cells in the hippocampus, the part of the brain that governs crystallized intelligence (Monje and Dietrich, 2012).

While we find the same effect for men and women, our results show significant differences between age cohorts. For the youngest cohorts in the sample — those born between 1976 and 1986 — we find no effect, while we find strong negative effects for older cohorts. 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. One potential explanation — consistent with the science literature — is that the effects of radiation only materialize at older age. Radiation induces a stochastic error in the regeneration of cells that increases with age (Rola et al., 2004). By contrast, the younger cohorts in our sample may be too young to experience cognitive decline due to radiation today, although they may experience these effects later in their lives. An additional interesting result is a stronger effect for people living in East

Germany in 1986 compared with those living in West Germany. One explanation for this finding is that the government in East Germany released little information about the disaster, whereby the population had fewer possibilities to engage in compensating behavior.

Our results imply that 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 the equivalent dose of two mammograms or 40% of a CT scan. According to our estimates, this dose reduces cognitive test scores by the equivalent of 0.05 school years. In the literature, similar effects have only been found for people exposed to radiation in the womb (Almond et al., 2009; Heiervang et al., 2010; Black et al., 2013) as well as for groups who are selectively exposed such as cancer patients who underwent radiotherapy (Hall et al., 2004; Pearce et al., 2012).² Our study shows that strong negative effects are present in the general population, suggesting that the human capital costs of man-made radiation are non-trivial. While medical procedures, nuclear energy and flying are undoubtedly beneficial, they come with the cost of higher radiation exposure of the population. For governments, there is scope to minimize these costs by reducing the number of unnecessary CT scans or investing in alternative sources of energy.

Chernobyl is one of the two largest nuclear disasters in history. With more data becoming available, a growing body of literature studies its long-run consequences. Besides the aforementioned studies on exposure *in-utero* (Almond et al., 2009; Heiervang et al., 2010), several studies provide evidence 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). Two studies document important negative effects of the fallout on human capital in Ukraine, the country where Chernobyl is located. Lehmann and Wadsworth (2011) find a negative effect of ground deposition on self-reported health and wages, while Danzer and Danzer (2016) find negative effects on subjective well-being. Our contribution to this literature is to show that the effect of the Chernobyl fallout on cognitive performance is by no means limited to people exposed in the womb or during early childhood. Our unique dataset allows us to measure cognitive skills at older ages, and our results show that radiation has significant negative effects even when people are first exposed as adults. In addition, with Germany being located more than 1,000km from Ukraine, our results point to significant external costs of nuclear power generation.

More broadly, our paper contributes to the literature on pollution and human capital. This literature has mainly documented two types of effects. First, exposure to pollution in the womb or during early childhood leads to lower birth weight, worse health, higher crime rates, as well as lower educational attainment, productivity and wages (see Almond and Currie (2011) and Graff Zivin and Neidell (2013) for reviews of the literature).³ Another strand of literature focuses on

²There is also evidence of a higher cancer risk among groups exposed to subclinical doses of radiation, such as flight attendants (Pukkala et al., 1995; Rafnsson et al., 2001; Haldorsen et al., 2001), or workers at nuclear power plants (Cardis et al., 2007).

³See, for example, Feigenbaum and Muller (2016), Aizer and Currie (2017) and Billings and Schnepel (2018) for the effect of lead exposure, Brinkel et al. (2009) and Chowdhury et al. (2015) for the health effects of contaminated

adolescents and adults and studies contemporaneous effects of pollution on worker productivity and test scores, whereby pollution and the outcomes are often measured on the same day.⁴ However, there is little evidence on the long-term impact of pollution when people are first exposed after early childhood. Knowing the size of this impact is important due to the number of affected people: at any given point in time, the number of adults and adolescents is a multiple of that of fetuses in the womb or young children. Our data allow us to estimate the effect of continuous exposure during adulthood over a 25-year period. The results show that the negative long-term effects of pollution are clearly not limited to exposure early in life.

The remainder of this paper is structured as follows. In Section 2, we provide the background of the Chernobyl nuclear disaster and the fallout in Germany. In Section 3, we summarize the medical literature on the effect of radiation and develop a conceptual framework that guides our analysis. In Section 4, we describe the dataset and provide descriptive statistics. Section 5 explains the identification strategy and discusses potential threats to identification. In Section 6, we present the main results and explore non-linear as well as heterogeneous effects. In Section 7, we carry out extensive robustness checks, before concluding in Section 8.

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 atmosphere, and — through strong east winds — carried all over Europe.⁵ 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 10. The fallout comprises four main isotopes, namely caesium-137 (Cs137), caesium-134 (Cs134), strontium-90 (Sr90)

water, Chay and Greenstone (2003) and Isen et al. (2017) for the effect of air pollution, and Deschenes et al. (2009) for the effect of climate change on birth weight.

⁴See, 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.

⁵Becquerel (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.

and iodine-131 (I131), which have half-lives of up to 30 years.⁶ 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 German government 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, 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/m² to 107 kBq/m², whereas soil is officially considered contaminated if the radioactivity exceeds 37 kBq/m² (UNSCEAR, 2000). The majority of the population lived in areas with radiation levels below 20 kBq/m², although a non-negligible number of people lived in areas with levels much higher than that.⁷

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 30 years in a highly-affected area is constantly — namely year after year until today — 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 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 six round trips Frankfurt-New York or 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.⁸ Due

⁶The 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 ¹³⁷Cs, ¹³⁴Cs, ⁹⁰Sr and ¹³¹I.

⁷See Figure 4b in Appendix B.

⁸The 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.,

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.

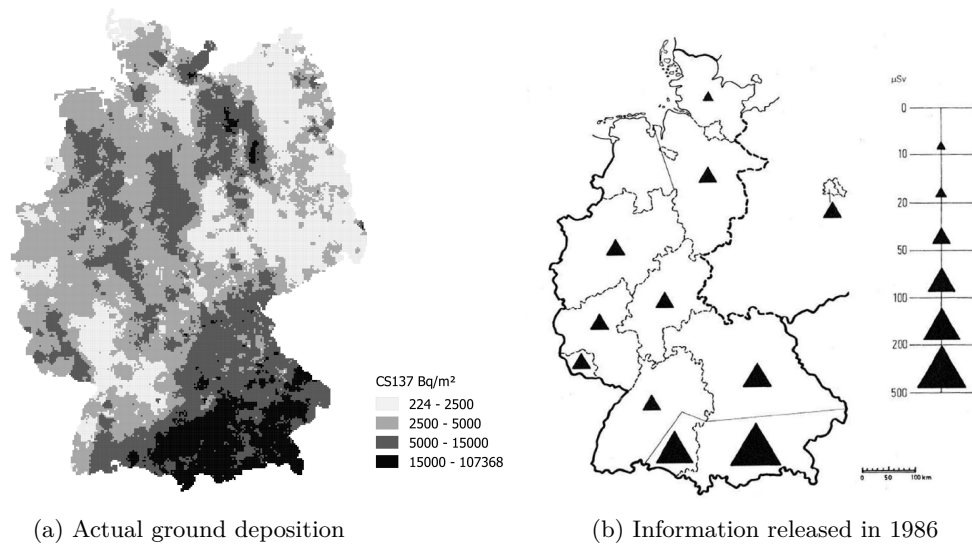


Figure 1: Ground contamination in 1986

Notes: These graphs display (a) the ground deposition of Cs137 in Bq/m² 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 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

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.

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 Protection⁹ 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

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

3.1 Exposure to radiation

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

In the northern hemisphere, the average yearly exposure to natural radiation is 2.4 mSv, of which 52% is through inhalation, 12% through ingestion, and 36% through terrestrial and cosmic radiation (UNSCEAR, 2008). The degree of exposure differs between people and depends on

⁹An Associate Society of the International Radiation Protection Association (IRPA)

their daily activities as well as their diet. For example, people who spend more time outdoors are more exposed to cosmic radiation than people who spend most of their time indoors, or people who are physically active — and therefore breathe more — have a higher exposure through inhalation.

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 damage the DNA, or indirectly damage the DNA when it ionizes water molecules in the cell. Radiobiology theory posits that a marginal increase in radioactivity linearly increases the probability that a cell is hit by an electron. A linear relationship emerges because during ionization the release of electrons follows a random process, such that each cell has an equal likelihood of being hit. Therefore, 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 is, the more likely it is that a critical mass of dysfunctional cells affects the functioning of organs and therefore leads to adverse health effects.

3.2 Conceptual framework

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

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

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. The channels shown in Equation (2) are firmly grounded in the scientific literature.

The effect of radiation on health The science literature provides ample evidence of negative health effects of radiation. These effects can either be *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. On the contrary, 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 such as flight attendants or workers at nuclear power plants, who receive an additional dose between 1 mSv and 2.5 mSv per year, are shown to have a higher cancer risk (UNSCEAR, 2008). In sum, this literature provides evidence that the direct effect on health is negative, $\frac{\partial H}{\partial R} < 0$.

The effect of radiation on cognition The effect of radiation on cognitive and neurodevelopmental functioning is an active research area in the sciences (OECD, 2016). Traditionally, the adult brain was considered resistant to radiation, as it was difficult to prove that brain cells regenerate (Deng et al., 2010). However, this view has been challenged in the last two decades. Cell regeneration has been found in the hippocampus, the part of the brain that governs several types of memory, in particular crystallized intelligence and learning (Squire, 2009; Supekar et al., 2013). Several studies show that a higher exposure to radiation slows down the regeneration of brain cells, which, in turn, impairs cognitive performance. Experimental research on animals finds that radiation reduces cell regeneration, which results in lower cognitive performance (Rola et al., 2004). Moreover, studies on humans show that exposure to low-dose radiation during medical treatments increases the risk of cognitive impairments several months

to years later (Hall et al., 2004; Douw et al., 2009; Monje and Dietrich, 2012). This evidence suggests that the direct effect on intelligence is negative or zero, $\frac{\partial I}{\partial R} \leq 0$

Behavioral responses 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. In light of this evidence, we expect the behavioral responses to counteract the effects on health and intelligence, i.e. $\frac{\partial B}{\partial R} \geq 0$.

3.3 Theoretical predictions

Equation (2) allows us to generate hypotheses about the sign of each channel. In light of the existing evidence, the direct effects are most likely negative, whereas the behavioral responses are most likely positive. Therefore, the sign of the overall effect is ambiguous. 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 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 with the interpretation of 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 Analysis

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, supervised and hosted by the Leibniz Institute for Educational Trajectories (LifBi, Blossfeld et al. (2011)). The NEPS 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 had been conducted by the Institute for Employment Research (IAB) in 2007 with originally 10,404 respondents.

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.¹⁰ A comparison of the ALWA subsample with the German Microcensus shows that, by and large, the sample is representative of the German population, although people with higher education and older people are slightly over-represented, whereas migrants are under-represented.

Besides including rich data on personal characteristics, employment histories and educational attainment, the NEPS SC6 offers two features that are key to our analysis. First, the survey includes standardized competence tests that allow us to measure cognitive skills along various dimensions for people who are between 24 and 58 years old. Second, the ALWA subsample includes detailed information on residential histories. In the initial survey wave, respondents were asked to provide monthly spell data on their municipality of residence since their birth. This allows us to link personal characteristics and cognitive test scores measured after 2009 with data on radiation levels in the person’s municipality of residence in May 1986.

4.2 Estimation sample

Our sample includes participants who were born before Chernobyl. 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 people born after April 1986. Therefore, our estimation sample only includes people born before the disaster.

¹⁰Of the 2,425 respondents who did not participate despite agreeing, 68% were unwilling, while 32% could not be contacted.

To reduce classification error, we drop respondents who moved in May 1986 (34 obs.), because 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 which 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, to address concerns about the representativeness of the estimation sample, we perform a series of balancing tests in Appendix D, which suggest that the missing information is unsystematic.

4.3 Cognitive 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. To make the estimates comparable across outcomes, we standardize the test scores to mean zero and standard deviation one. Moreover, given that the test scores measure different aspects of 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.¹¹

4.4 Municipality-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 highly 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 a 3×3 km-grid-cell level. The data were compiled by the BfS following a comprehensive measurement program that was rolled out between 1986 and 1989. At 3,474 temporary measurement points, gamma spectrometers measured the gamma radiation from Cs137 in the ground.¹² Based on the decay of Cs137, the BfS backdated all measures to May 1986. To compute the radiation in each grid

¹¹To construct the index, we first sum over all eight standardized test scores, and then standardize this sum to mean zero and standard deviation one.

¹²See Figure 8a in Appendix F for the location of measurement points in West and East Germany.

cell, the BfS used the average inverse-distance-weighted radiation of the four nearest measurement points. This is the only available dataset with detailed regional information on radiation. 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, but 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.¹³ 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 C.3 we perform robustness checks that show that the results are not driven by the linkage procedure.

Additional data We supplement our dataset with municipality-level data on precipitation, altitude and population density. The precipitation data is collected daily by 544 stations across the country.¹⁴ To link the precipitation with the survey data, we compute for each municipality centroid the inverse-distance-weighted average rainfall based on the four closest measuring stations. To measure the average altitude of a municipality — an important determinant for rainfall as well as radiation — we use data provided by the Service Center of the Federal Government for Geo-Information and Geodesy. From the same source, we also obtained data on population size for all municipalities in Germany. The earliest available geo-coded data on the population size of municipalities is from 1997.¹⁵ In line with LIfBi's data protection policy, the municipality information has been merged by the research data center. To ensure that municipalities cannot be identified based on this information, we rounded the environmental data.¹⁶

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.

¹³The 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.

¹⁴The data are provided by the German Meteorological Service. See Figure 8 for the location of the measuring points.

¹⁵The NEPS uses the official municipality codes of 2013. For municipalities that have been merged between 1997 and 2013, we use the sum of the merged municipalities.

¹⁶Cs137 is rounded to the nearest 0.1kBq, altitude rounded to the nearest 50m, population rounded to the nearest 5,000, precipitation rounded to the first 0.1mm/m².

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 before had much lower educational attainment than their children. Over half of all respondents have parents with no more than 9 years of schooling.

The fourth set of statistics describe the cognitive test scores. Two features are important 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. The statistics were computed across individual observations in the estimation sample. The mean ground deposition of Cs137 in May 1986 amounts to $5.18kBq/m^2$. The standard deviation, which is larger than the mean, points to a significant variation in ground deposition across Germany.¹⁷ We also compute the average radiation from Cs137 between 1986 and 2010. Given that radiation data is only available for 1986, we compute this average for each municipality across all years, correcting for the decay of Cs137.¹⁸ While this measure does not account for the washing out of Cs137 into deeper layers of soil, it serves as a proxy for the average potential exposure in an area. The level of precipitation represents the average rainfall in May in the five years preceding the Chernobyl disaster, i.e. 1981-1985. The average person lived in 1986 in a medium-sized municipality with 282,000 inhabitants, although municipality sizes vary between 5,000 and over 3 million.

¹⁷See Appendix B for an illustration of the distribution of the ground deposition across municipalities.

¹⁸For each year t since 1986, we apply the decay formula for Cs137, $Cs137_t = e^{-0.024t}$.

Table 1: Descriptive statistics of the main variables

	Mean	(SD)	min	max	N
A. Individual-level data					
<i>Personal characteristics</i>					
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
<i>Educational attainment in April 1986</i>					
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, no degree	0.43	0.49	0.00	1.00	4440
already attained lower secondary, secondary	0.33	0.47	0.00	1.00	4440
already attained upper secondary	0.10	0.31	0.00	1.00	4440
already attained tertiary	0.01	0.09	0.00	1.00	4440
0.00	0.00	0.00	0.00	0.00	4440
<i>Highest parental education</i>					
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
<i>Test Scores</i>					
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 comprehension	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 (average 1981-1985)	3.09	0.84	1.30	8.00	4440
Altitude in meters	201.59	176.69	0.00	850.00	4440
Minimum altitude in meters in county	138.73	139.78	-1.00	660.00	4440
Population/1000	281.67	676.43	5.00	3420.00	4440

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

5 Empirical Strategy

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

$$y_{ims} = \alpha + \beta CS137_{ms}^{86} + \mathbf{X}'_{im}\boldsymbol{\gamma} + \delta_s + \varepsilon_{ims}. \quad (3)$$

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 \mathbf{X}_{im} controls for pre-treatment characteristics of individuals and municipalities as well as some 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, origin, parental education, education in 1986, and employment status in 1986. It also includes municipality characteristics, namely the average daily rainfall between 1981 and 1985, altitude and log population.¹⁹ 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.²⁰ In some specifications, we also control for state fixed effects, δ_s .²¹ 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. Therefore, β is to be interpreted as a reduced-form coefficient. A higher initial radiation in a municipality 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.

¹⁹The 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 both 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 G provides further details on the control variables.

²⁰See Appendix A.2 for a description of the survey design.

²¹In 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 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.

5.2 Identification

Identifying assumption The parameter β can only be interpreted as causal if the ground deposition of Cs137 is uncorrelated with any determinants of test scores that are not controlled for in Equation (4), i.e. $E(\varepsilon_{im} \times CS137_{im} | \mathbf{X}_{im}, \delta_s) = 0$. Empirical studies on the impact of pollution typically face two major threats to identification, namely anticipation and residential sorting (Graff Zivin and Neidell, 2013). In the case of Chernobyl, anticipation can be excluded because people could neither anticipate the disaster nor foresee which area would be more contaminated. In fact, most of the German population only learned about the disaster after the radioactive rain had fallen. Importantly, anticipation — an ex-ante change in behavior in expectation of contamination — is not to be confused with ex-post avoidance behavior. It is possible that people responded to high exposure with changes in their lifestyle — for example, by changing their diet or exercise habits — or move to a less contaminated area. While affecting people’s effective exposure, behavioral responses do not invalidate the identifying assumption; rather, they represent a channel through which radiation affects test scores.

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

Balancing tests on the raw data point to a correlation between (pre-1986) residential sorting and the level of fallout received in 1986. In Columns (1)-(3) of Table 2, we compare the pre-treatment characteristics of people who lived in 1986 in municipalities with below- and above-median ground deposition of Cs137. While the sample is balanced on some characteristics, there is evident sorting on skills. 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, tend to be less populated, and have more rainfall. In other words, less skilled Germans tend to live in more rural areas, and rural areas received a greater level of nuclear fallout after the Chernobyl disaster due to their altitude and rainfall levels.

In Columns (4)-(9), we test whether the sample is balanced conditional on controls. 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 states. Conditional on these controls, the sample is balanced on all observable characteristics. Therefore, our preferred specification will include

controls for municipality-level characteristics as well as state fixed effects.

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

We address these challenges using several methods, namely balancing tests, selection on observables (Oster, 2017), as well as a series of robustness checks. We discuss the implications of these tests along with the main estimation results in the next section, and provide further details in the appendix. However, for the interpretation of the results to follow, we work with the maintained assumption that, conditional on state fixed effects and controls for municipality characteristics, the ground deposition of Cs137 is uncorrelated with any personal determinants of cognitive test scores.

5.3 Inference

To account for potential cross-sectional dependence of the error terms, we cluster the standard errors at the county level. 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 an adjustment of standard errors (Benjamini and Hochberg, 1995).

Table 2: Sample description: Balancing table

	Raw data			Control municipality charac.			State FE, Municipality charac.		
	Below (1)	Mean Above (2)	Diff (2)-(1) (3)	Below (4)	Mean Above (5)	Diff (5)-(4) (6)	Mean (7)	Diff (8)-(7) (9)	
A. Individual characteristics									
Age in 1986	18.947 (0.178)	19.137 (0.170)	0.190 (0.246)	-0.242 (0.178)	0.227 (0.169)	0.470* (0.245)	-0.149 (0.178)	0.140 (0.169)	0.289 (0.245)
Female	0.517 (0.011)	0.499 (0.010)	-0.018 (0.015)	0.007 (0.011)	-0.007 (0.015)	-0.014 (0.015)	0.003 (0.011)	-0.003 (0.010)	-0.006 (0.015)
Native speaker	0.974 (0.003)	0.980 (0.003)	0.006 (0.004)	0.001 (0.003)	0.001 (0.003)	0.003 (0.004)	-0.001 (0.003)	0.001 (0.003)	0.001 (0.004)
Employed in April 1986	0.363 (0.010)	0.351 (0.010)	-0.012 (0.014)	0.005 (0.010)	-0.004 (0.010)	-0.009 (0.014)	-0.002 (0.010)	0.002 (0.010)	0.004 (0.014)
Unemployed in April 1986	0.014 (0.003)	0.013 (0.002)	-0.002 (0.003)	0.000 (0.003)	0.000 (0.002)	0.000 (0.003)	0.002 (0.003)	-0.002 (0.002)	-0.004 (0.003)
If employed: Qualified or highly qualified job before May 1986	0.519 (0.017)	0.524 (0.017)	0.005 (0.025)	0.003 (0.017)	-0.003 (0.017)	-0.007 (0.024)	0.001 (0.017)	-0.001 (0.017)	-0.001 (0.024)
Children before 1986	0.178 (0.008)	0.154 (0.008)	-0.024** (0.011)	0.008 (0.008)	-0.008 (0.008)	-0.016 (0.011)	-0.005 (0.008)	0.005 (0.007)	0.010 (0.011)
Older siblings	0.521 (0.011)	0.546 (0.011)	0.024 (0.016)	0.006 (0.011)	0.006 (0.011)	0.012 (0.011)	-0.007 (0.011)	0.007 (0.011)	0.014 (0.015)
Smoke before 1986	0.346 (0.010)	0.369 (0.010)	0.023 (0.014)	0.019 (0.010)	0.019 (0.010)	0.039** (0.014)	-0.011 (0.010)	0.010 (0.010)	0.021 (0.014)
<i>Educational attainment in April 1986</i>									
Lower secondary and secondary	0.032 (0.004)	0.043 (0.004)	0.011** (0.006)	-0.004 (0.004)	0.004 (0.004)	0.009 (0.006)	-0.001 (0.004)	0.001 (0.004)	0.002 (0.006)
Upper secondary	0.280 (0.010)	0.289 (0.009)	0.009 (0.014)	-0.002 (0.010)	0.002 (0.009)	0.004 (0.014)	0.000 (0.010)	-0.000 (0.009)	-0.001 (0.013)
Tertiary	0.137 (0.007)	0.116 (0.007)	-0.021** (0.010)	0.006 (0.007)	-0.006 (0.007)	-0.012 (0.010)	-0.001 (0.007)	0.001 (0.007)	0.002 (0.010)
In school or college education	0.424 (0.011)	0.434 (0.010)	0.010 (0.015)	-0.004 (0.011)	0.004 (0.010)	0.009 (0.015)	-0.001 (0.011)	0.001 (0.010)	0.002 (0.015)
In education, already attained lower secondary and secondary	0.103 (0.007)	0.104 (0.006)	0.001 (0.009)	-0.009 (0.006)	0.009 (0.006)	0.018** (0.009)	-0.007 (0.006)	0.007 (0.006)	0.014 (0.009)
In education, already attained upper secondary	0.010 (0.002)	0.006 (0.002)	-0.005* (0.003)	0.001 (0.002)	-0.001 (0.002)	-0.003 (0.003)	0.002 (0.002)	-0.001 (0.002)	-0.003 (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 (0.007)	0.119 (0.007)	-0.008 (0.010)	0.005 (0.007)	-0.005 (0.007)	-0.010 (0.010)	0.003 (0.007)	-0.002 (0.007)	-0.005 (0.010)
<i>Highest parental education</i>									
Lower secondary education	0.467 (0.011)	0.575 (0.010)	0.108*** (0.015)	-0.030 (0.011)	0.029 (0.010)	0.059** (0.015)	-0.010 (0.011)	0.009 (0.010)	0.019 (0.015)
Secondary education	0.288 (0.010)	0.249 (0.009)	-0.039*** (0.013)	0.006 (0.010)	-0.006 (0.009)	-0.012 (0.013)	-0.000 (0.010)	0.000 (0.009)	0.001 (0.013)
Upper secondary	0.246 (0.009)	0.176 (0.008)	-0.069*** (0.012)	0.024 (0.007)	-0.023 (0.008)	-0.047** (0.012)	0.010 (0.009)	-0.009 (0.008)	-0.019 (0.012)
B. Municipality characteristics									
Caesium137 kBq/m ² (01. May 1986)	2.283 (0.014)	7.905 (0.150)	5.622*** (0.150)	2.283 (0.014)	7.905 (0.150)	5.622*** (0.150)	2.283 (0.014)	7.905 (0.150)	5.622*** (0.150)
Altitude in meter	141.419 (2.852)	258.079 (4.048)	116.660*** (4.952)	141.419 (2.852)	258.079 (4.048)	116.660*** (4.952)	141.419 (2.852)	258.079 (4.048)	116.660*** (4.952)
Minimum altitude in meter in county	88.299 (1.705)	186.072 (3.459)	97.773*** (3.856)	88.299 (1.705)	186.072 (3.459)	97.773*** (3.856)	88.299 (1.705)	186.072 (3.459)	97.773*** (3.856)
Population	438.286 (19.409)	134.627 (5.994)	-303.659*** (20.313)	438.286 (19.409)	134.627 (5.994)	-303.659*** (20.313)	438.286 (19.409)	134.627 (5.994)	-303.659*** (20.313)
Precipitation in mm/m ²	2.911 (0.021)	3.264 (0.014)	0.353*** (0.025)	2.911 (0.021)	3.264 (0.014)	0.353*** (0.025)	2.911 (0.021)	3.264 (0.014)	0.353*** (0.025)
GDR	0.289 (0.010)	0.082 (0.006)	-0.207*** (0.011)	0.289 (0.010)	0.082 (0.006)	-0.207*** (0.011)	0.289 (0.010)	0.082 (0.006)	-0.207*** (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. 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 the estimation results for the impact of radiation exposure on cognitive test scores. We begin with reduced-form regressions of test scores in 2010 on the initial exposure in 1986. Given that people were continuously exposed for 25 years, we also estimate the effect of the the average level of radiation between 1986 and 2010 on test scores. We further explore the non-linear dose-response relationship as well as heterogeneous effects across demographic groups. Finally, we assess the role of three behavioral responses, namely investment in education, labor supply and migration.

6.1 Main results

Reduced-form results Table 3 reports the reduced-form estimates. Each entry represents an estimate for β from a separate regression of the outcomes listed on the left on the ground deposition of Cs137 in kBq/m^2 in May 1986 and the controls listed at the bottom. The table comprises two sets of results; in Columns (1)-(4) the outcomes are the raw test scores, whereas in Columns (5)-(8), the outcomes are standardized to mean zero and standard deviation one. In each set, we begin with a bivariate regression and gradually add control variables. Given the results from the balancing test in Section 5, our preferred specification includes controls for municipality characteristics as well as state fixed effects.

In the bivariate regressions in Column (1), some estimates are positive and others negative. In Column (2), most estimates become positive after we control for individual characteristics. This is unsurprising given that i) people with less education lived in areas that received more fallout, and ii) education presumably has a positive effect on cognitive test scores. In Columns (3) and (4), once we control for municipality characteristics and state fixed effects, all coefficients are negative, and, in Column (4), four out of eight coefficients are statistically significant. Perhaps surprisingly, the estimates become larger once we control for state fixed effects. One explanation for this is the differences in education systems across German regions. Students in the southern states of Bavaria and Baden-Württemberg usually achieve the highest test scores in standardized tests such as the Programme for International Student Assessment (PISA).²² At the same time, these regions received considerably more fallout than other parts of the country. By controlling for state fixed effects, we eliminate the between-state variation in fallout and test scores. In light of the balancing tests in Table 2, the specification in Column (4) is our preferred one, given that the state fixed effects eliminate the — potentially confounding — correlation between individual characteristics and the level of fallout.

Based on the results in Column (4), an increase in ground deposition by one standard deviation ($5.86 kBq/m^2$) reduces math scores by 0.3 points, which is 2.5% of the mean. For reading, the corresponding effect is -0.58 points, or 2.1% of the mean reading score. The effect sizes for the other outcomes range between -0.2% and -1.02% of the mean test scores.

²²See Baumert et al. (2002).

To make the estimates comparable between test scores, in Columns (5)-(8), we report the estimated effects on standardized outcomes. The interpretation of the coefficients is such that for an increase in the initial ground deposition by $1 \text{ kBq}/\text{m}^2$, the test score changes on average by $100 \times \hat{\beta}$ percent of a standard deviation. The point estimates suggest that radiation has the largest impact on reading scores, followed by math, listening comprehension and reading speed, while it has smaller impacts on ICT skills, scientific literacy, perceptual speed and reasoning.

In Panel B, we additionally report the estimated impacts of radiation on a standardized cognitive skills index, as well as indices for skills relying on crystallized and fluid intelligence. For all three indices, the effects are large and statistically significant. For a one-unit increase in ground deposition, the standardized index decreases by 0.8% of a standard deviation. The effect on skills based on crystallized intelligence is larger (-0.9% of a standard deviation) than the one on skills based on fluid intelligence (-0.6% of a standard deviation). To evaluate the economic significance of our estimates, it is useful to consider the estimated impact of a one-standard-deviation increase in ground deposition in 1986. This decreases reading scores 25 years later by 8%, math scores by 6.3%, listening comprehension scores by 5.1% and reading speed by 4.5%, and ICT scores by 2.8% of the respective standard deviation. The impact on the overall cognitive skills index is -4.5% of a standard deviation. These results point to a high economic significance of the impact of radiation on cognitive skills.

The effect of average exposure 1986-2010 The reduced-form 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 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.²³

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 4, 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

²³Even in medical research, it is difficult to precisely measure the effective dose. Rather, the effective dose is estimated by simulation (McCullough and Schueler, 2000).

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

	Non-standardized outcomes				Standardized outcomes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Individual test scores								
Math	0.014	0.010	-0.041**	-0.053***	0.003	0.002	-0.009**	-0.011***
<i>mean = 11.316</i>	(0.018)	(0.014)	(0.016)	(0.015)	(0.004)	(0.003)	(0.003)	(0.003)
Reading	-0.011	0.009	-0.056*	-0.102***	-0.001	0.001	-0.008*	-0.014***
<i>mean = 27.065</i>	(0.042)	(0.033)	(0.029)	(0.037)	(0.006)	(0.004)	(0.004)	(0.005)
Listening comprehension	-0.025	-0.024	-0.051*	-0.076**	-0.003	-0.003	-0.006*	-0.009**
<i>mean = 75.821</i>	(0.030)	(0.026)	(0.027)	(0.032)	(0.004)	(0.003)	(0.003)	(0.004)
ICT	0.004	0.016	-0.024	-0.074	0.000	0.001	-0.002	-0.005
<i>mean = 41.199</i>	(0.030)	(0.029)	(0.037)	(0.045)	(0.002)	(0.002)	(0.003)	(0.003)
Scientific literacy	0.005	0.013	-0.011	-0.021	0.001	0.002	-0.002	-0.004
<i>mean = 18.997</i>	(0.015)	(0.013)	(0.015)	(0.016)	(0.003)	(0.002)	(0.003)	(0.003)
Reasoning	0.005	0.004	0.002	-0.003	0.002	0.001	0.001	-0.001
<i>mean = 8.939</i>	(0.008)	(0.007)	(0.010)	(0.010)	(0.003)	(0.003)	(0.004)	(0.004)
Reading speed	-0.007	0.007	-0.036	-0.065**	-0.001	0.001	-0.004	-0.008**
<i>mean = 38.187</i>	(0.028)	(0.023)	(0.033)	(0.030)	(0.003)	(0.003)	(0.004)	(0.004)
Perceptual speed	0.023	0.024	-0.014	-0.032	0.003	0.003	-0.002	-0.004
<i>mean = 34.679</i>	(0.023)	(0.018)	(0.025)	(0.025)	(0.003)	(0.002)	(0.003)	(0.003)
B. Indices								
Cognitive skill index					0.001	0.002	-0.004	-0.008***
					(0.003)	(0.003)	(0.003)	(0.003)
Crystallized intelligence index					0.000	0.002	-0.005*	-0.009***
					(0.003)	(0.003)	(0.003)	(0.003)
Fluid intelligence index					0.002	0.003	-0.002	-0.006*
					(0.003)	(0.003)	(0.003)	(0.003)
<i>Controls:</i>								
Individual characteristics	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Municipality characteristics	No	No	Yes	Yes	No	No	Yes	Yes
State FE	No	No	No	Yes	No	No	No	Yes

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

(1) reproduces the main regression results with the ground deposition in 1986 as the regressor. 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 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 Column (4), we address two 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. We address both problems by instrumenting the average ground deposition between 1986 and 2010 with the *initial* ground deposition in a person's municipality of residence in 1986. Unsurprisingly, the first stage is strong with an F-statistic greater than 4,500. 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 Column (4) are considerably larger than the OLS estimates in Column (3). For example, the effect on the cognitive skills index is 90% larger. An increase in the average ground deposition by one standard deviation reduces cognitive test scores by 6.1% of a standard deviation.

Discussion of the main results The estimates presented in Tables 3 and 4 suggest that the radiation induced by Chernobyl had significant negative long-term effects on cognitive perfor-

Table 4: The Effect of Average Exposure, 1986-2010

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)
A. Individual test scores				
Math	-0.011*** (0.003)	-0.015*** (0.004)	-0.014** (0.006)	-0.026*** (0.009)
Reading	-0.014*** (0.005)	-0.018*** (0.007)	-0.013* (0.007)	-0.034*** (0.010)
Listening comprehension	-0.009** (0.004)	-0.013** (0.005)	-0.012* (0.007)	-0.023** (0.010)
ICT	-0.005 (0.003)	-0.007 (0.004)	-0.003 (0.005)	-0.013 (0.008)
Scientific literacy	-0.004 (0.003)	-0.005 (0.004)	-0.002 (0.005)	-0.009 (0.008)
Reasoning	-0.001 (0.004)	-0.001 (0.006)	0.000 (0.007)	-0.003 (0.010)
Reading speed	-0.008** (0.004)	-0.010** (0.005)	-0.020*** (0.006)	-0.019** (0.009)
Perceptual speed	-0.004 (0.003)	-0.005 (0.004)	-0.006 (0.006)	-0.010 (0.009)
B. Indices				
Cognitive skill index	-0.008*** (0.003)	-0.011*** (0.004)	-0.010* (0.005)	-0.019** (0.008)
Crystallized intelligence index	-0.009*** (0.003)	-0.012*** (0.004)	-0.008* (0.005)	-0.020*** (0.008)
Fluid intelligence index	-0.006* (0.003)	-0.008* (0.004)	-0.011 (0.007)	-0.014* (0.008)
First stage:				
Cs137 kBq/m ² in 1986				0.425*** (0.006)
F statistic				4555.589
<i>Controls:</i>				
Individual characteristics	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Mean (Cs137)	5.18	3.89	2.95	2.95
SD (Cs137)	5.87	4.41	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 (8) 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. The mean and standard deviation of Cs137 refer to the regressor used in each column. 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: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

mance. People who have been living in areas with higher contamination for 25 years perform significantly worse than those living in less contaminated area. The fact that the local ground deposition was mainly driven by rainfall in a short time window and that personal characteristics are uncorrelated with the level of ground deposition lends strong support to the effect being causal.

To assess the magnitude of the effect, it is useful to compare the radiation effective dose from the Chernobyl fallout to that of other man-made sources of radiation. For example, during a Frankfurt-New York return flight, passengers receive a dose of 0.1mSv, mainly through cosmic radiation. The average woman who undergoes a mammogram receives a dose of 0.4mSv, whereas a person who has a CT scan receives a dose of 2mSv. By comparison, one standard deviation in the initial ground deposition of Cs137 in kBq corresponds to a cumulative effective dose of 0.7mSv.²⁴ Because the effective dose is measured as the extent of cell destruction in organs and human tissue, these doses are directly comparable even if they come from different sources. An effective dose of 1mSv does the same damage to the human body regardless whether it is received at once or through continuous exposure over a long time.

Relative to these benchmarks, our estimates point to substantial long-term effects of radiation. A person who lived continuously in a place with a one-standard-deviation higher ground deposition received an additional dose equivalent to 1.75 mammograms, 35% of a CT scan, or seven additional flights Frankfurt-New York and back. Our estimates predict that over 25 years, this person’s cognitive performance declines by 5-6% of a standard deviation. With one percent of a standard deviation in the cognitive skills test being equivalent to the human capital of a school year, this means that receiving this additional radiation dose reduces a person’s human capital by the equivalent of 5-6% of a school year.²⁵

6.2 Non-linear and heterogeneous effects

Non-linear dose-response relationship. In Table 5, we analyze if there is a non-linear dose-response relationship between radiation exposure and cognitive test scores. In each regression, the outcome is the cognitive skills index. For comparison, Column (1) reproduces the linear estimate reported in Column (8) of Table 3.

The estimates in Columns (2) and (4) provide little evidence in favor of a non-linear relationship. In Column (2), we impose a quadratic relationship, but find no significant coefficient for the quadratic term. In Column (4), we estimate a spline regression by interacting the ground deposition with a binary indicator that equals unity if a person lived in 1986 in an area with above-median ground deposition. While the point estimate is larger for people living in areas

²⁴We base this figure on the estimate of the BfS that the mean effective dose until 2010 is 0.6mSv. Given that a standard deviation in initial Cs137 (5.87) is larger than the mean (5.18), we scale the mean effective dose, $0.6 * \frac{5.87}{5.18} \approx 0.7$.

²⁵The 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.

with above-median ground deposition, the coefficient is statistically insignificant, such that a linear relationship cannot be rejected. In Column (3), we impose a log-linear relationship, for which we find a large and statistically significant coefficient. For a one-standard-deviation increase in the log ground deposition ($sd=0.72$), we find a decrease in cognitive test scores by 5.6% of a standard deviation, which is similar to the estimate from the linear level-level model in Column (1).

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

Table 5: Non-linear effects

	(1)	(2)	(3)	(4)
CS137 kBq/m ²	-0.008*** (0.003)	-0.014** (0.006)		-0.032 (0.033)
CS137 kBq/m ² × CS137 kBq/m ²		0.000 (0.000)		
ln(CS137 Bq/m ²)			-0.083*** (0.029)	
CS137 kBq/m ² × above median				-0.118 (0.090)
<i>Controls:</i>				
Individual 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 1986 in an area with an above-median ground deposition. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

Heterogeneous effects In Table 6, 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.²⁶ 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 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 exposure to radiation — those on brain cells and vital organs — are more likely to manifest themselves in older age. However, because the youngest cohort was only 25 years old when they took the cognitive skills tests, we cannot observe what their test scores would be when the same cohort is in their mid-50s. Another potential explanation is that parents with young kids in 1986 particularly tried to shield their children away, thereby reducing the 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

²⁶We 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.

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

Table 6: Heterogenous effects

	(1)	(2)	(3)	(4)
CS137 kBq/m ² × male	-0.008*			
	(0.003)			
CS137 kBq/m ² × female	-0.008**			
	(0.003)			
CS137 kBq/m ² × Age in 1986(0-10)		0.003		
		(0.005)		
CS137 kBq/m ² × Age in 1986(10-20)		-0.018***		
		(0.007)		
CS137 kBq/m ² × Age in 1986(>20)		-0.007**		
		(0.003)		
CS137 kBq/m ² × Parent(above secondary education)			-0.004	
			(0.003)	
CS137 kBq/m ² × Parent(below secondary education)			-0.010***	
			(0.003)	
CS137 kBq/m ² × West Germany				-0.008***
				(0.003)
CS137 kBq/m ² × East Germany				-0.016
				(0.015)
<i>Controls:</i>				
Individual 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	0.22	0.22	0.22

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

6.3 Evidence on behavioral responses

In Table 7, 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, diets or exercise habits.

In the first panel of Table 7, 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. 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. In Appendix D, we additionally consider the impact of radiation on cohort-specific mortality, but find no evidence of any effect.

Table 7: 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	3.938	(7.903)
Employed in 2000	0.000	(0.001)
Employed in 2005	-0.001	(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	-1.453	(0.586)**
<i>Controls:</i>		
Individual characteristics	Yes	
Municipality characteristics	Yes	
State FE	Yes	

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

7 Robustness checks

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

7.1 Addressing unobserved heterogeneity

The balancing tests in Section 5 show that, conditional on state fixed effects and municipality-level controls for altitude, rainfall and population size, the sample is balanced with respect to a large set of observable characteristics. Nonetheless, we are unable to observe all determinants of test scores. Because these unobserved determinants could be correlated with radiation exposure, it is possible that the omission of unobservables leads to biased estimates.

7.1.1 Selection on observables

We first assess the potential influence of unobservable characteristics on our estimates. Oster (2017) provides a method that quantifies the importance of unobservable characteristics based on the selection on observable characteristics. From the difference between the treatment effects in regressions with and without controls, it can be inferred how strong the selection on unobservables has to be to obtain a treatment effect of zero. This strength is summarized by the selection parameter δ . In Appendix C.1, we apply this procedure to all eleven outcomes. For all regressions, we obtain negative values for δ , which means that, to explain our estimated treatment effect, selection on unobservables, would have to go in the opposite direction than selection on observables. Therefore, it is unlikely that our estimates are driven by selection on unobservables.

7.1.2 Regressions with grid-level fixed effects

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

As an alternative, we lay a grid of $120 \times 120 km$ cells over the map of Germany and estimate a model with grid cell fixed effects. This approach exploits variation within grid cells, thereby comparing people with the same characteristics living in the same grid cell but are differentially exposed to radiation. To account for the arbitrary locus of the grid, we perform 500 replications

of the same procedure, whereby in each replication we shift the grid by random distances north-south and east-west.²⁷

Figure 2 plots the empirical distributions of the estimates for all eleven outcomes. The results confirm our baseline findings in Section 6. For some outcomes such as math, reading and listening, the estimates are large and the entire distribution of estimates lies below zero. For other outcomes, such as reasoning or reading speed, the estimates are centered around zero. In Appendix C.2, we additionally report the average p-value for each outcome, which provides a robustness check to our inference.

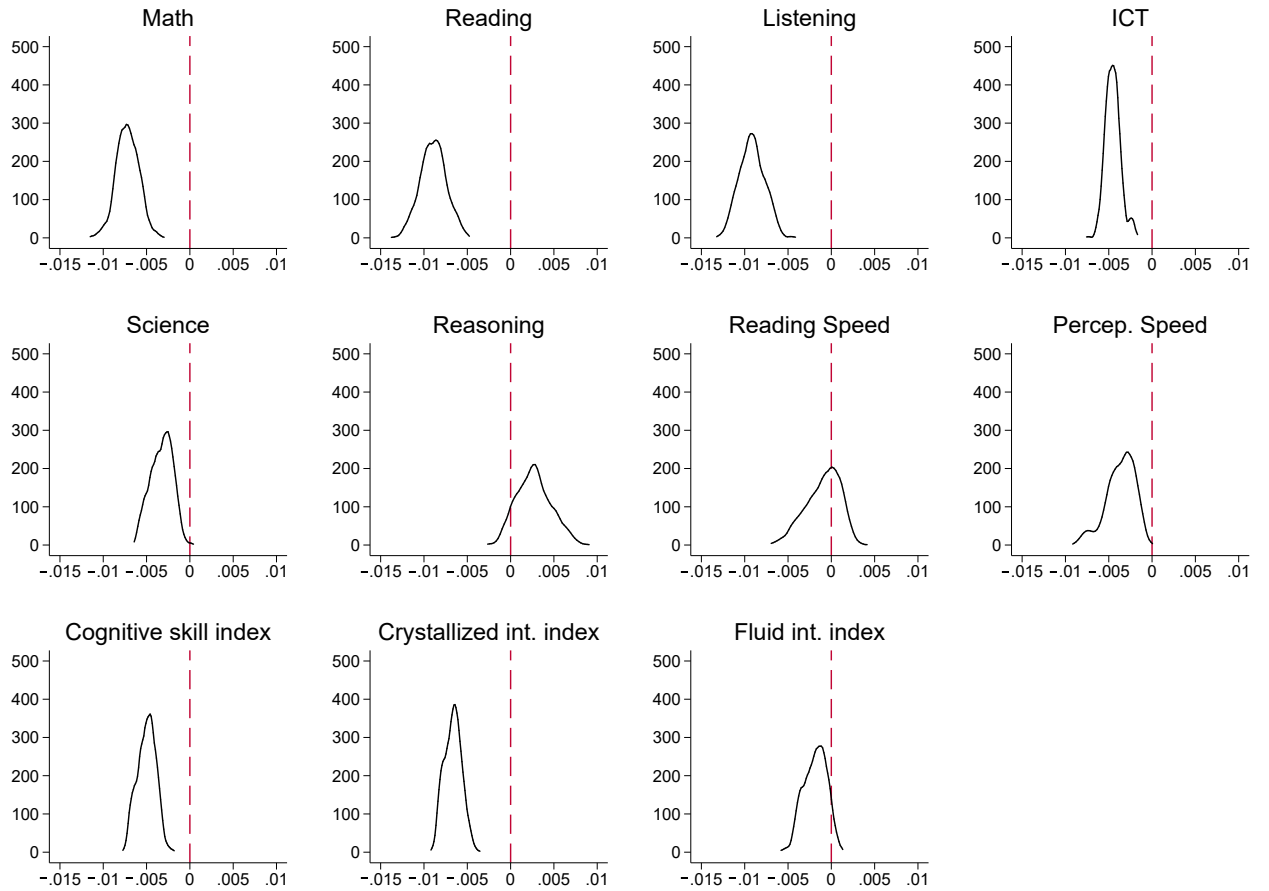


Figure 2: Empirical distribution of fixed effect estimates with random grid cells

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

²⁷This approach follows Barsbai et al. (2017). We are grateful to Andreas Steinmayr for sharing his code.

7.2 Addressing measurement error

An important potential source of measurement error is the linkage of radiation data via the centroid of a municipality. The geographic center of a municipality can be in a different location compared to the population center, such that the average person is in fact exposed to a different level of radiation than the one at the geographic center. To address this concern, we perform a series of eight robustness checks, whereby in each robustness check we link the data based on different geographic reference points. For example, we link the data based on the population center or the point with the highest population density, or compute the population-weighted average radiation based on the radiation data at a $3 \times 3km$ grid-cell level. The results, shown in Appendix C.3, turn out to be robust to the linkage procedure, suggesting our choice of linkage procedure does not drive the results.

7.3 Selective attrition

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

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

Constructing a standardized index in the presence of attrition. One of the outcomes reported with our baseline results is a standardized index of all eight cognitive test scores. However, as explained in Section 4, due to the survey design not every person in the sample completed every test. While there appears to be no selective attrition from each test, attrition presents a challenge for the construction of the standardized index and the interpretation of the

results. For the estimation in Section 6, we constructed the standardized index based on all test scores a person obtained; if someone completed all eight tests, the index is the average score of these eight tests; if someone completed one test, the index is this one test score.

To test for the robustness of our results to the construction of the index in the presence of attrition, we estimate separate regressions with samples restricted to people who completed at least a number of tests. The results are reported in Appendix D.3. For example, out of 4,400 observations, 1,029 completed all eight tests, 2,159 completed at least seven tests, and 3,207 completed at least five tests. The estimates are larger the more tests people completed, ranging from -0.013 for those with all eight tests to -0.008 for those with at least one test. All results are statistically significant at the 5%-level or lower. This suggests that, if anything, by using all available test scores to construct the index, we under-estimate the true effect.

7.4 Inference

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

The statistical literature proposes two approaches to overcome the multiple testing problem.²⁸ One is to keep the number of tests constant but adjust the p-values and t-statistics of each test to account for the correlations between outcomes. The other is to leave the p-values and t-statistics unadjusted but reduce the number of hypotheses to one by using a standardized index as outcome.

In Appendix E.2, we report the results of both approaches. To adjust the p-values, we use the step-down procedure developed by Benjamini and Hochberg (1995), which first ranks the p-values of all hypotheses tests from smallest to lowest, and adjusts each p-value depending on its rank position. The smallest p-value is adjusted the most — it is multiplied by the number of hypotheses tested — while the largest p-value is not adjusted at all.²⁹ After the correction of the p-values, the effects on three outcomes, namely math, reading and listening, remain statistically significant at the 5%-level, and the one on reading speed at the 10%-level.

As an alternative, we perform a summary index test whereby all outcomes are summarized

²⁸See Anderson (2008) for a discussion and applications.

²⁹An often-used alternative is the Bonferroni correction, which simply multiplies each p-value by the number of hypothesis tests. However, as shown by Benjamini and Hochberg (1995), the Bonferroni correction is very conservative and leads to a severe under-rejection of the null hypothesis, while the step-down procedure provides a more accurate adjustment of p-values.

in a standardized index. Following O’Brien (1984) and Anderson (2008), we construct the index by standardizing each outcome to mean zero and standard deviation one and building for each observation a weighted sum of the standardized outcomes. The weights are constructed such that outcomes that are highly correlated — and therefore add little new information to the index — receive lower weight than outcomes with a low correlation that add more new information.³⁰ Using the weighted index, the estimated effects are similar to those on the standardized index reported in Section 6, in terms of both point estimates and standard errors.

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

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

Cluster bootstrap-t procedure (Cameron et al., 2008) An alternative to a permutation test is to cluster the standard errors at the appropriate level and compute cluster-robust standard errors. In our case, the most appropriate level of clustering appears to be the state level. Education policies in Germany are set at the state level, such that outcomes or error terms may be correlated across people living in the same state. However, with a total of twelve federal states — eleven West German states plus the GDR — the number of clusters is too small to apply conventional cluster-robust standard errors, whose asymptotic validity relies on the number of clusters going to infinity. Cameron et al. (2008) propose a cluster bootstrap procedure that provide an asymptotic refinement to conventional cluster-robust standard errors and reduce the

³⁰Technically, the weights are constructed from the row sums of the covariance matrix of all outcomes, whereby the elements on the diagonal are set to zero. An outcome that has a higher correlation with all others has a higher row sum. This inverse of this sum is used as weight. See Appendix E.2 for further details.

problem of over-rejection of the null hypothesis. In Appendix E.3, we estimate the standard errors clustered at the state level using Cameron et al. (2008)’s wild cluster bootstrap. The standard errors are larger than those obtained from conventional clustering at the county level, but most results remain statistically significant at the 10% or 5%-level.

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 the effect to be 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 adults. For policy-makers, these results are important for at least two reasons. First, they point to substantial external costs of nuclear power generation. Chernobyl is over 1,000km away from the German border and yet 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 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. 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

- Aizer, A. and Currie, J. (2017). Lead and Juvenile Delinquency: New Evidence from Linked Birth, School and Juvenile Detention Records. Working Paper 23392, National Bureau of Economic Research.
- Alinaghizadeh, H., Wålinder, R., Vingård, E. and Tondel, M. (2016). Total cancer incidence in relation to (137)cs fallout in the most contaminated counties in sweden after the chernobyl nuclear power plant accident: a register-based study. *BMJ Open* 6: e011924.
- Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives* 25: 153–172.
- Almond, D., Edlund, L. and Palme, M. (2009). Chernobyl’s subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly Journal of Economics* 124: 1729–1772.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association* 103: 1481–1495.
- Antoni, M., Drasch, K., Kleinert, C., Matthes, B., Ruland, M. and Trahms, A. (2011). Working and learning in a changing world: Part I: Overview of the study-March 2011 (Second, updated version). Tech. rep., Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].
- 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. Tech. rep., Infas - Institut für Angewandte Sozialwissenschaften.
- 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.
- Barsbai, T., Rapoport, H., Steinmayr, A. and Trebesch, C. (2017). The effect of labor migration on the diffusion of democracy: Evidence from a former soviet republic. *American Economic Journal: Applied Economics* 9: 36–69.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. and Weiss, M. (2002). PISA 2000 - Die Länder der Bundesrepublik Deutschland im Vergleich. Tech. rep., Max-Planck-Institut für Bildungsforschung, Berlin.

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57: 289–300.
- Berendes, K., Weinert, S., Zimmermann, S. and Artelt, C. (2013). Assessing language indicators across the lifespan within the german national educational panel study (neps)/die erfassung von sprachindikatoren über die lebensspanne im nationalen bildungspanel. *Journal for Educational Research Online* 5: 15.
- Billings, S. and Schnepel, K. (2018). Life after lead: Effects of early interventions for children exposed to lead. *American Economic Journal: Applied Economics* .
- 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* .
- Blomstrand, M., Holmberg, E., Åberg, M. A. I., Lundell, M., Björk-Eriksson, T., Karlsson, P. and Blomgren, K. (2014). No clinically relevant effect on cognitive outcomes after low-dose radiation to the infant brain: A population-based cohort study in sweden. *Acta Oncologica (Stockholm, Sweden)* 53: 1143–1150.
- Blossfeld, H.-P., Rossbach, H.-G. and Maurice, J. von (2011). Education as a lifelong process - the german national educational panel study (neps). *Zeitschrift für Erziehungswissenschaft* 14 [Special Issue].
- Bolus, N. E. (2013). Ncrp report 160 and what it means for medical imaging and nuclear medicine. *Journal of Nuclear Medicine Technology* 41: 255–260.
- 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, PMID: 18046031.
- Brinkel, J., Khan, M. H. and Kraemer, A. (2009). A systematic review of arsenic exposure and its social and mental health effects with special reference to bangladesh. *International Journal of Environmental Research and Public Health* 6: 1609–1619.
- Brunner, M., Lang, F. R. and Lüdtke, O. (2014). Erfassung der fluiden kognitiven leistungsfähigkeit über die lebensspanne im rahmen der national educational panel study: Expertise [assessment of fluid cognitive skills over the life span in the national educational panel study: Expertise]. *NEPS Working Paper* .

- Bundesamt für Strahlenschutz (2016). Der Reaktorunfall 1986 in Tschernobyl. Tech. rep., Bundesamt für Strahlenschutz.
- Bundesregierung (1986-1991). Bericht der Bundesregierung über Umweltradioaktivität und Strahlenbelastung. Tech. rep., Deutscher Bundestag.
- Bunzl, K., Schimmack, W., Krouglov, S. and Alexakhin, R. (1995). Changes with time in the migration of radiocesium in the soil, as observed near chernobyl and in germany, 1986–1994. *Science of the Total Environment* 175: 49–56.
- 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.
- Cardis, E., Vrijheid, M., Blettner, M., Gilbert, E., Hakama, M., Hill, C., Howe, G., Kaldor, J., Muirhead, C. R., Schubauer-Berigani, M., Yoshimura, T., Bermann, F., Cowper, G., Fix, J., Hacker, C., Heinmiller, B., Marshall, M., Thierry-Chef, I., Utterback, D., Ahn, Y. O., Amoros, E., Ashmore, P., Auvinen, A., Bae, J. M., Bernar, J., Biau, A., Combalot, E., Deboodt, P., Sacristan, A. D., Eklöf, M., Engels, H., Engholm, G., Gulis, G., Habib, R. R., Holan, K., Hyvonen, H., Kerekes, A., Kurtinaitis, J., Malke, H., Martuzzi, M., Mastauskas, A., Monnet, A., Moser, M., Pearce, M. S., Richardson, D. B., Rodriguez-Artalejo, F., Rogel, A., Tardy, H., Telle-Lamberton, M., Turai, I., Usel, M. and Veress, K. (2007). The 15-country collaborative study of cancer risk among radiation workers in the nuclear industry: Estimates of radiation-related cancer risks. *Radiation Research* 167: 396–416.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*, 35. Elsevier.
- Chay, K. Y. and Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession*. *The Quarterly Journal of Economics* 118: 1121–1167.
- Chowdhury, S., Krause, A. and Zimmermann, K. F. (2015). Arsenic contamination of drinking water and mental health. *IZA Discussion Paper* 9400.
- 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.

- Deng, W., Aimone, J. B. and Gage, F. H. (2010). New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nature Reviews. Neuroscience* 11: 339.
- Deschenes, O., Greenstone, M. and Guryan, J. (2009). Climate change and birth weight. *American Economic Review* 99: 211–17.
- 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.
- Feigenbaum, J. J. and Muller, C. (2016). Lead exposure and violent crime in the early twentieth century. *Explorations in Economic History* 62: 51 – 86.
- Fielitz, U. and Richter, K. (2013). Bundesweiter Überblick über die Radiocäsiumkontamination von Wildschweinen. Tech. rep., Bundesamt für Strahlenschutz (BfS).
- Gehrer, K., Zimmermann, S., Artelt, C. and Weinert, S. (2012). The Assessment of Reading Competence (Including Sample Items For Grade 5 and 9. Tech. rep., Leibniz Institute for Educational Trajectories.
- Gehrer, K., Zimmermann, S., Artelt, C. and Weinert, S. (2013). Neps framework for assessing reading competence and results from an adult pilot study/neps-rahmenkonzeption zur messung von lesekompetenz und resultate einer pilotstudie mit erwachsenen. *Journal for Educational Research Online* 5: 50.
- 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.
- Haberkorn, K. and Pohl, S. (2013). Cognitive Basic Skills (Non Verbal) Data in the Scientific Use File. Tech. rep., University of Bamberg, National Educational Panel Study (NEPS),.

- Hachenberger, C., Trugenberger-Schnabel, A., Löbke-Reinl, A. and Peter, J. (2017). Umweltradioaktivität und Strahlenbelastung—Jahresbericht 2015. Tech. rep., Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU).
- Haldorsen, T., Reitan, J. B. and Tveten, U. (2001). Cancer incidence among norwegian airline cabin attendants. *International Journal of Epidemiology* 30: 825–830.
- Hall, P., Adami, H.-O., Trichopoulos, D., Pedersen, N. L., Laggiou, 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., Dahlgard, 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: n/a–n/a, rG1001.
- Ihme, J. M., Senkbeil, M. and Wittwer, J. (2015). The NEPS ICT Literacy Framework and Item Examples. Tech. rep., Leibniz Institute for Educational Trajectories.
- 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.
- Isen, A., Rossin-Slater, M. and Walker, W. R. (2017). Every breath you take—every dollar you’ll make: The long-term consequences of the clean air act of 1970. *Journal of Political Economy* 125: 848–902.
- 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.
- Nature (1992). Thyroid cancer after chernobyl. *Nature* 359: 21, copyright - Copyright Macmillan Journals Ltd. Sep 3, 1992; Zuletzt aktualisiert - 2017-10-31; CODEN - NATUAS.
- Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E. and Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan/modellierung und erfassung mathematischer kompetenz über die lebensspanne. *Journal for Educational Research Online* 5: 80.
- O’Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* 4: 1079–1087.
- OECD (2016). Radiological Protection Science and Application. Tech. rep., OECD.
- Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* forthcoming.
- 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 Bevölkerung auf die Ereignisse in Tschernobyl - Ergebnisse einer Befragung. Tech. Rep. Juel-Spez-0400, Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag, Jülich.
- 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.
- Pukkala, E., Auvinen, A. and Wahlberg, G. (1995). Incidence of cancer among finnish airline cabin attendants, 1967-92. *BMJ* 311: 649–652.
- Rafnsson, V., Tulinius, H., Jónasson, J. G. and Hrafnkelsson, J. (2001). Risk of breast cancer in female flight attendants: A population-based study (iceland). *Cancer Causes & Control* 12: 95–101.

- Reimer, M. and Matthes, B. (2007). Collecting event histories with true tales: 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). Mathematical Competence: Framework and Exemplary Test Items. Tech. rep., Leibniz Institute for Educational Trajectories.
- 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. Tech. rep., United Nations Scientific Committee on the Effects of Atomic Radiation and others.
- UNSCEAR (2000). Sources and effects of ionizing radiation. Tech. rep., United Nations Scientific Committee on the Effects of Atomic Radiation and others.
- UNSCEAR (2008). Sources and effects of ionizing radiation. Tech. rep., United Nations Scientific Committee on the Effects of Atomic Radiation and others.
- van den Ham, A.-K., Ehmke, T., Hahnand, I., Wagner, H. and Schöps, K. (2016). Mathematische und naturwissenschaftliche Kompetenz in PISA?, im IQB-Ländervergleich und in der National Educational Panel Study (NEPS) – Vergleich der Rahmenkonzepte und der dimensional Struktur der Testinstrumente. In *Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments*. Bundesministerium für Bildung und Forschung, 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., Kuchida, D., Müller, M. et al. (1986). Ergebnisse von radioaktivitätsmessungen nach dem reaktorunfall in tschernobyl. *ISH-99, Institut für Strahlenhygiene des Bundesgesundheitsamtes, Neuherberg* .
- Winkelmann, I., Haubelt, R., Neumann, P. and Fields, D. (1989). Radionuclide deposition and exposure in the Federal Republic of Germany after the Chernobyl accident. Tech. rep., Oak Ridge National Lab., TN (USA).
- Yasunari, T. J., Stohl, A., Hayano, R. S., Burkhart, J. F., Eckhardt, S. and Yasunari, T. (2011). Cesium-137 deposition and contamination of japanese soils due to the fukushima nuclear accident. *Proceedings of the National Academy of Sciences* 108: 19530–19534.
- Zimmermann, S., Artelt, C. and Weinert, S. (2014). The Assessment of Reading Speed in Adults and First-Year Students. Tech. rep., Leibnitz Institute for Educational Trajectories.

Online Appendices

(Not for publication)

A Data description	47
A.1 Sampling in the ALWA subsample	47
A.2 Competence tests	48
A.3 Participation in the competence tests	50
B Further Descriptive Statistics	52
C Addressing unobserved heterogeneity and measurement error	52
C.1 Selection on observables, Oster (2017)	52
C.2 Regressions with grid-level fixed effects	53
C.3 Robustness to different data linkage procedures	53
D Testing for systematic sample selection	58
D.1 Selective mortality	58
D.2 Design-based attrition	58
D.3 The cognitive skills index with non-participation	59
E Inference	63
E.1 Randomization inference	63
E.2 Multiple hypothesis testing	63
E.3 Cluster bootstrap-t procedure (Cameron et al., 2008)	66
F Geographic information	68
G Control variables	72

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 assess competence development across 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 categorized 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 Peabody Picture Vocabulary Test (PPVT) which is used in several large surveys such as the British Cohort Study, the European Child Care and Education 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 procedure 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 procedure 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 3c, some only started in the second and some few only in the third period.

As shown in figure 3b 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 3a 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

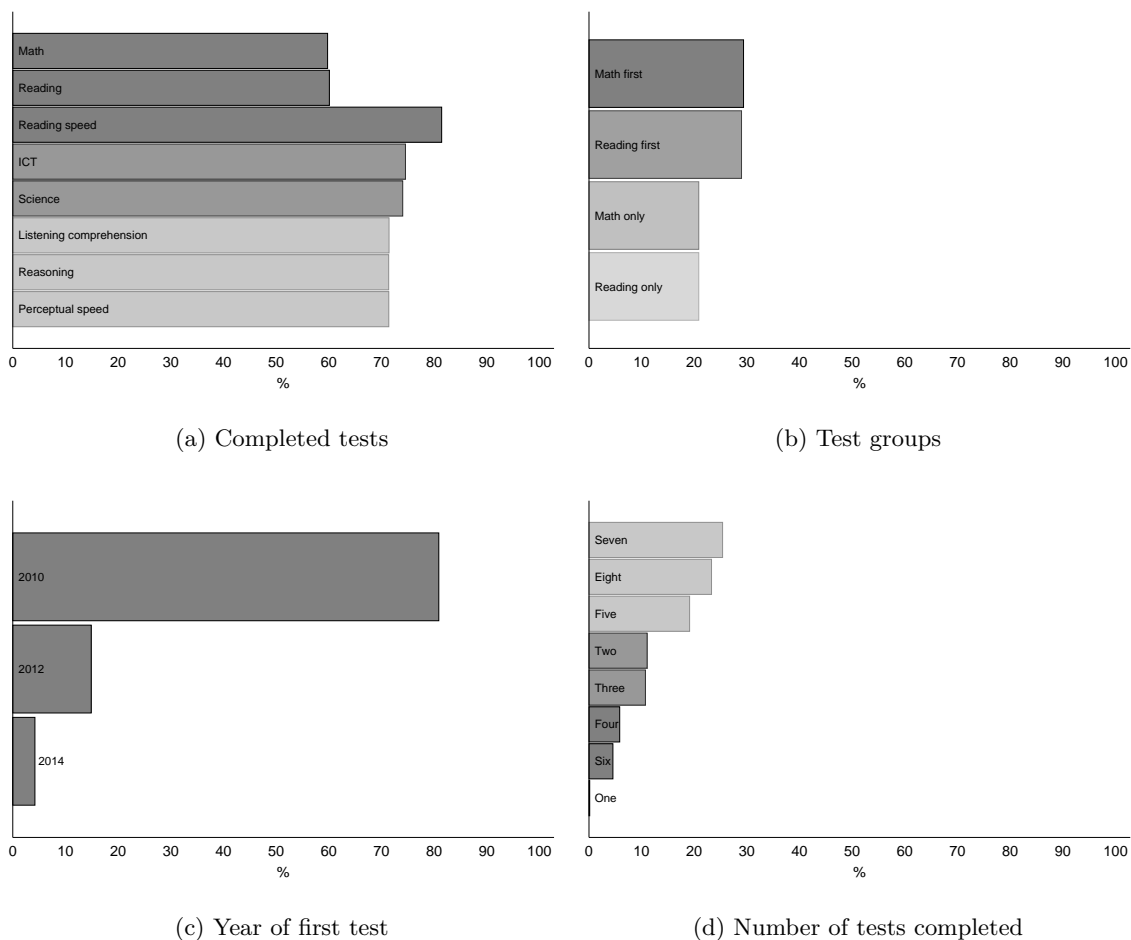


Figure 3: Participation in cognitive tests

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

participants performed at least one test. Figure 3d 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 D.2, we test whether the non-participation in the competence tests is systematically linked to the level of radiation, which is not the case.

B Further Descriptive Statistics

Figure 4 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.

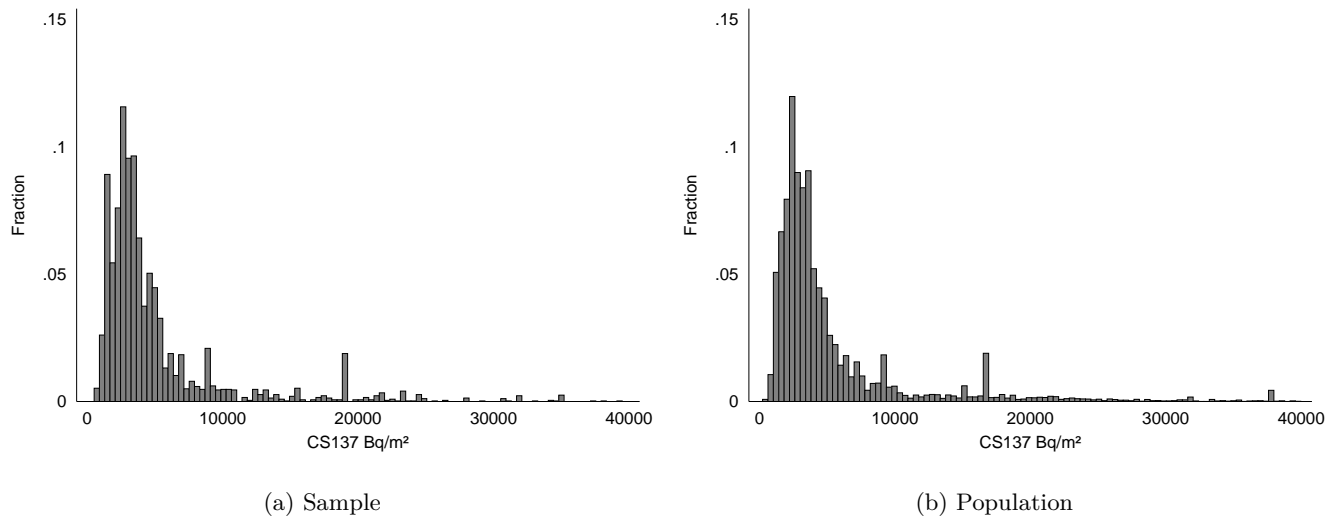


Figure 4: Variation in the ground deposition of Cs137 in May 1986

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.

C Addressing unobserved heterogeneity and measurement error

C.1 Selection on observables, Oster (2017)

One of the potential threats to identification is the presence of unobserved heterogeneity between people exposed to high versus low levels of fallout. While the balancing tests in Table 2 suggests that the sample, conditional on controls and fixed effects, is balanced on observable characteristics, we cannot be sure that the same holds for unobservable characteristics. To assess the importance of unobservable characteristics in explaining our results, we apply the bounding method of Oster (2017). This method infers the selection on unobservables from the selection on observables. In doing so, the method provides a statistic δ that shows how strong the selection on unobservables would have to be for the estimate to become zero. At a value of $\delta = 1$, the selection on unobservables would have to be as strong as the selection on observables and — in

addition — both selections would have to go in the same direction. A negative δ , by contrast, means that if the observables are positively correlated with the treatment, the unobservables have to be negatively correlated.

Table 8 displays the estimates for δ . In Column (1), δ is estimated based on a comparison of a bivariate regression — the test score regressed on Cs137 — with a multivariate regression that includes all controls and fixed effects.³¹ All estimates of δ are negative and range between -0.428 and -0.065, indicating that, if anything, the selection of unobservables would have to have the opposite sign to bring the estimate for β to zero. In Column (2), we isolate the variation to within states, which is congruent with our identifying variation. In that case, most estimates of δ are large in absolute value and, as before, all estimates are negative. In light of these results, it is unrealistic that our estimates are driven by unobserved heterogeneity. The balancing tests in Table 2 point to a negative correlation between observable characteristics and the treatment — people with a lower education and less educated parents had a higher exposure to radiation. It is difficult to think of factors that would plausibly be positively correlated with radiation and — in addition — this correlation would have to be between 36% and 125% of the correlation of observables with Cs137.³²

In addition to the estimates of δ , we also calculate the adjusted R^2 to see if the model fit improves as more characteristics are added. As shown in Columns (1) and (2) of Table 9, radiation by itself explains little of the total variation in individual test scores. However, the model fit is greatly improved if individual characteristics are added. Additional controls for municipality characteristics and state fixed effects (Columns 3 and 4), in turn, add little explanatory power. However, these controls are important to ensure the balancedness of the sample between people living in high- vs. low-exposure areas.

C.2 Regressions with grid-level fixed effects

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

C.3 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

³¹We assume that $R^{max} = 1$, which means that the observable and unobservable characteristics can fully explain test scores.

³²The small proportionality factor for reasoning ($\delta = -0.035$) is an outlier here.

Table 8: Selection on unobservables

	(1)	(2)
A. Individual test scores		
Math	-0.167	-0.494
Reading	-0.365	-1.043
Listening comprehension	-1.086	-0.369
ICT	-0.219	-1.152
Scientific literacy	-0.143	-0.663
Reasoning	0.055	-0.034
Reading speed	-0.163	-1.278
Perceptual speed	-0.098	-0.442
B. Indices		
Cognitive skill index	-0.224	-0.672
Crystallized intelligence index	-0.238	-0.796
Fluid intelligence index	-0.110	-0.355
<i>Controls:</i>		
State FE	No	Yes

Notes: This table displays the estimates for δ , the proportionality factor between selection on observable and unobservable characteristics. The calculation is based on the method proposed by Oster (2017) that compares the treatment effect of a restricted regression with and an unrestricted regression without controls. In Column (1), we condition in both regressions on state fixed effects, whereas in Column (2), the unrestricted regression is a bivariate regression of the dependent variable on the ground deposition.

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 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 11, we re-estimate the baseline model from Table 3, Column (8), with regressors based on different data linkages. The results in Table 11 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 municipality³³

³³We computed the population center as the balancing point of a municipality based on night light data from

Table 9: Adjusted R^2 for main results.

	(1)	(2)	(3)	(4)
A. Individual test scores				
Math	0.00	0.19	0.20	0.21
Reading	0.00	0.17	0.19	0.21
Listening comprehension	0.00	0.10	0.12	0.12
ICT	0.00	0.21	0.22	0.22
Scientific literacy	0.00	0.19	0.19	0.19
Reasoning	0.00	0.12	0.13	0.14
Reading speed	0.00	0.08	0.09	0.12
Perceptual speed	0.00	0.26	0.27	0.27
B. Indices				
Cognitive skill index	0.00	0.19	0.21	0.22
Crystallized intelligence index	0.00	0.18	0.20	0.21
Fluid intelligence index	0.00	0.17	0.18	0.20
<i>Controls:</i>				
Individual characteristics	No	Yes	Yes	Yes
Municipality characteristics	No	No	Yes	Yes
State FE	No	No	No	Yes

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

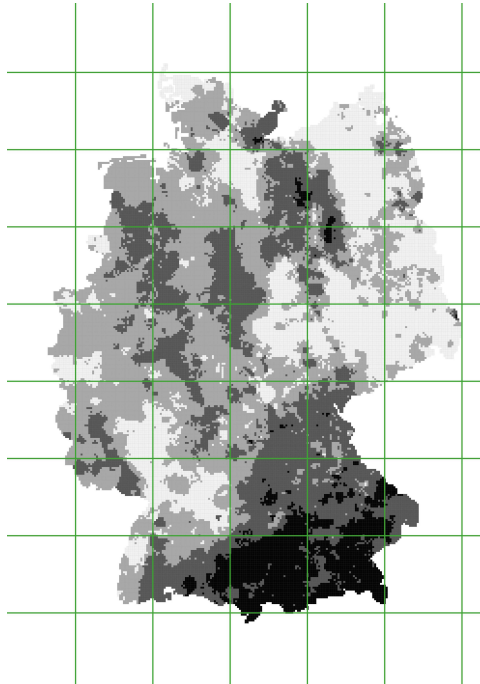


Figure 5: 120km \times 120km grid cells and regional variation in caesium-137 ground contamination in May 1986

1996 provided by NASA.

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

	Average coefficient	Average p-value	Share of p-values with		
			$p < 0.1$	$p < 0.05$	$p < 0.01$
	(1)	(2)	(3)	(4)	(5)
A. Individual test scores					
Math	-0.007	0.082	0.710	0.444	0.044
Reading	-0.009	0.088	0.640	0.192	0.000
Listening comprehension	-0.009	0.042	0.946	0.700	0.074
ICT	-0.005	0.227	0.010	0.000	0.000
Scientific literacy	-0.003	0.346	0.056	0.000	0.000
Reasoning	0.003	0.641	0.000	0.000	0.000
Reading speed	-0.001	0.697	0.010	0.000	0.000
Perceptual speed	-0.004	0.332	0.050	0.000	0.000
B. Indices					
Cognitive skill index	-0.004	0.112	0.462	0.174	0.008
Crystallized intelligence index	-0.007	0.052	0.906	0.562	0.014
Fluid intelligence index	-0.002	0.665	0.000	0.000	0.000
<i>Controls:</i>					
Individual characteristics	Yes				
Municipality characteristics	Yes				
State FE	Yes				

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

- Column (4): based on the radiation at the closest measuring point, linked via the population center of a municipality
- Column (5): based on the inverse-distance-weighted average radiation of the four closest measuring points, linked via the population mode of a municipality³⁴
- 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 municipality³⁵

³⁴We take as population mode the point in a municipality with the highest light intensity in 1996.

³⁵The averages in Columns (7) and (8) were computed based on the 3x3km grid-level data. To construct the population weights, we used night light data from 1996.

Table 11: Robustness to the data linkage procedure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Individual test scores								
Math	-0.011*** (0.003)	-0.008*** (0.002)	-0.012*** (0.005)	-0.009*** (0.003)	-0.014*** (0.005)	-0.012*** (0.004)	-0.013*** (0.004)	-0.012*** (0.004)
Reading	-0.014*** (0.005)	-0.008** (0.004)	-0.017*** (0.005)	-0.010** (0.004)	-0.017*** (0.005)	-0.007 (0.004)	-0.016*** (0.005)	-0.015*** (0.005)
Listening comprehension	-0.009** (0.004)	-0.005* (0.003)	-0.009** (0.004)	-0.006* (0.003)	-0.007 (0.005)	-0.005 (0.004)	-0.012*** (0.005)	-0.012*** (0.004)
ICT	-0.006* (0.003)	-0.002 (0.002)	-0.007* (0.004)	-0.004 (0.003)	-0.003 (0.004)	0.001 (0.003)	-0.006* (0.004)	-0.006* (0.004)
Scientific literacy	-0.004 (0.003)	-0.002 (0.002)	-0.005 (0.004)	-0.003 (0.003)	-0.003 (0.004)	-0.002 (0.003)	-0.005 (0.004)	-0.004 (0.004)
Reasoning	-0.001 (0.004)	0.001 (0.003)	0.001 (0.005)	0.002 (0.004)	0.002 (0.005)	0.001 (0.004)	-0.002 (0.005)	-0.002 (0.005)
Reading speed	-0.008** (0.003)	-0.005* (0.002)	-0.010*** (0.04)	-0.004 (0.004)	-0.011** (0.005)	-0.004 (0.004)	-0.010** (0.004)	-0.009** (0.004)
Perceptual speed	-0.004 (0.003)	0.002 (0.002)	-0.005 (0.004)	-0.003 (0.003)	-0.004 (0.004)	-0.002 (0.003)	-0.005 (0.004)	-0.005 (0.004)
B. Indices								
Cognitive skill index	-0.008*** (0.003)	-0.005** (0.002)	-0.010*** (0.003)	-0.005** (0.003)	-0.009** (0.004)	-0.004 (0.003)	-0.010*** (0.004)	-0.010*** (0.004)
Crystallized intelligence index	0.009*** (0.003)	-0.005** (0.002)	-0.011*** (0.003)	-0.007** (0.003)	-0.010** (0.004)	-0.004 (0.003)	-0.011*** (0.004)	-0.010*** (0.004)
Fluid intelligence index	-0.006* (0.003)	-0.003 (0.002)	-0.007* (0.004)	-0.002 (0.003)	-0.006 (0.004)	-0.002 (0.003)	-0.007* (0.004)	-0.007* (0.004)
<i>Controls:</i>								
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table displays the estimation results whereby the regressor has been constructed with different data linkage procedures. See text for a description of the linkage procedures. The controls are the same as in Table 3, Column (8). Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

D Testing for systematic sample selection

D.1 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).³⁶ We run the following regression:

$$m_{crt} = \alpha + \rho_{rt} Cs137_{crt} + \mathbf{X}'_{cs} \boldsymbol{\kappa} + \delta_s + \varepsilon_{crt}. \quad (4)$$

The number of deaths m_{crt} 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, \mathbf{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_{crt} 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.

D.2 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 12 shows, there is no evidence of systematic attrition or non-response once we add appropriate controls.

In Table 13, 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 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.

³⁶Such detailed data is only available from 1995 onwards

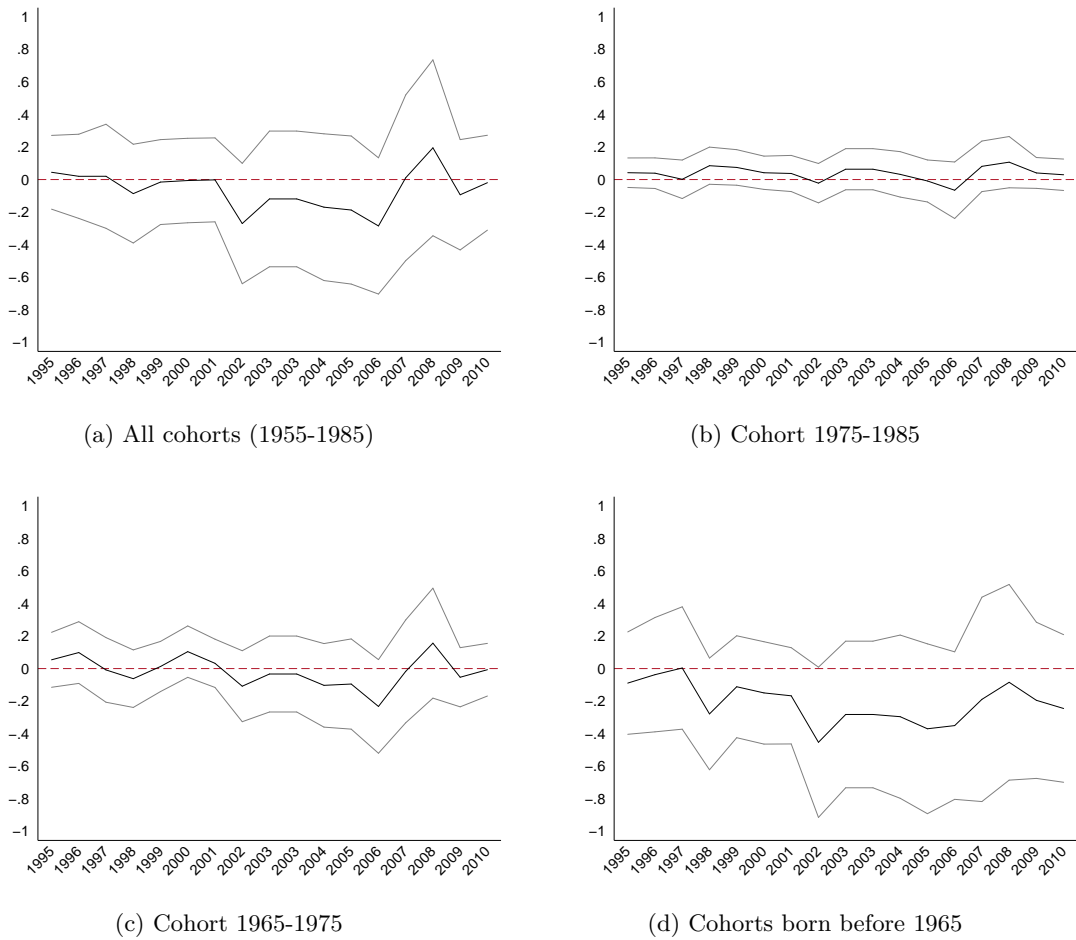


Figure 6: Radiation exposure and mortality.

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

D.3 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 14 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

Table 12: Selection into competence tests

	(1)	(2)	(3)	(4)
Math	-0.002 (0.002)	-0.003 (0.002)	-0.001 (0.002)	0.000 (0.000)
Reading	-0.000 (0.001)	-0.000 (0.002)	0.001 (0.002)	0.000 (0.000)
Listening comprehension	0.000 (0.001)	0.002* (0.001)	0.003* (0.001)	0.002 (0.001)
ICT	0.002** (0.001)	0.002 (0.001)	0.003** (0.001)	0.002* (0.001)
Scientific literacy	0.002** (0.001)	0.002 (0.001)	0.003** (0.001)	0.002 (0.001)
Reasoning	0.000 (0.001)	0.002* (0.001)	0.003* (0.001)	0.002 (0.001)
Reading speed	-0.001 (0.001)	-0.002 (0.002)	-0.001 (0.001)	-0.000 (0.000)
Perceptual speed	0.000 (0.001)	0.002* (0.001)	0.003* (0.001)	0.002 (0.001)
<i>Controls:</i>				
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: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

based on all respondents leads to smaller estimates than calculating the index based on those who completed seven or eight tests.

Table 13: 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
<i>Controls:</i>			
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 14: The cognitive skills index with different definitions

	(Coef.)	(N)
All eight tests	-0.014*** (0.005)	1034
At least seven tests	-0.012*** (0.004)	2159
At least six tests	-0.013*** (0.004)	2360
At least five tests	-0.010*** (0.004)	3207
At least four tests	-0.010*** (0.003)	3466
At least three tests	-0.010*** (0.003)	3942
At least two tests	-0.008*** (0.003)	4430
At least one test	-0.008*** (0.003)	4440
<i>Controls:</i>		
Individual characteristics	Yes	
Municipality characteristics	Yes	
State FE	Yes	

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

E Inference

E.1 Randomization inference

To assess the reliability of 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 7a displays the placebo distribution of 10,000 estimates with randomization across all observations, which allows us to assess the inference in a model without state fixed effects (Table 3 Column (3)). If the relationship was pure noise, a point estimate at least as extreme as -0.008 would be very unlikely to occur. In fact, in 10,000 replications, such a result only occurred once. The distribution in Figure 7b corresponds to the estimations with state fixed effects presented in Table 3 Column (4). In this test, we randomize the regressor within states and otherwise follow the same procedure as before. Again, an estimate at least as extreme as our point estimate of -0.008 would be very unlikely to occur by chance. In 10,000 estimations, it occurred 26 times, i.e. in 0.026% of all cases. This corresponds to an empirical p-value in a one-sided test of $p = 0.00026$.

Figure 7c 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 7d 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.

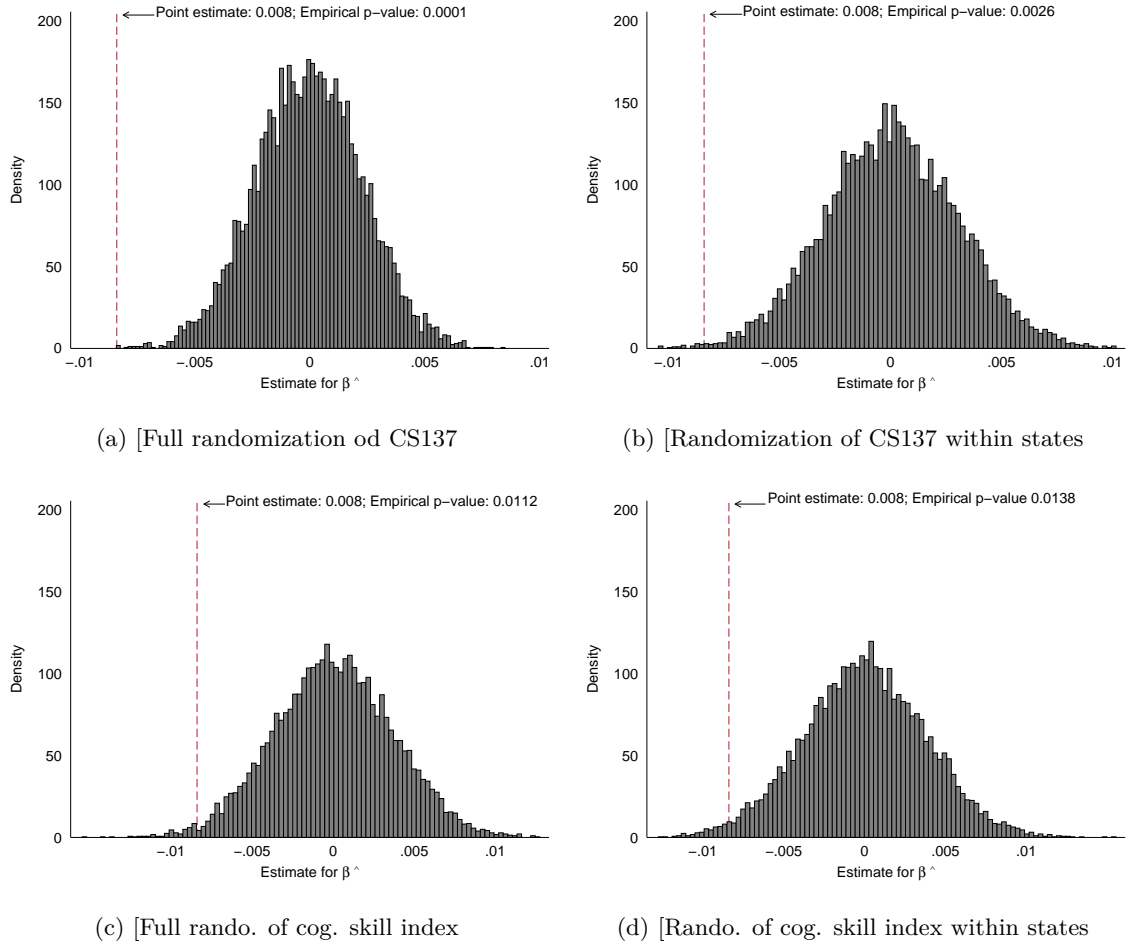


Figure 7: Randomization inference

Notes: This figure displays the empirical distributions of the estimates for $\hat{\beta}$ under the null hypothesis of no treatment effect based on 10,000 replications. In each replication, we randomize the ground deposition while keeping the outcome — the standardized index — and all other regressors fixed. In Panel (a), the treatment is randomized across all observations; in Panel (b), it is randomized across observations within states. 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

$$q = \frac{pm}{m - (i - 1)} \quad (5)$$

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 15 displays the p-values and q-values for all eight outcomes.³⁷ After the adjustment, three coefficients remain statistically significant at the 5%-level and one (reading speed) at the 10%-level.

Table 15: Q-values (p-values adjusted by step-down approach)

	(1)	(2)
	p-values	q-values
Math	0.001	0.006
Reading	0.005	0.022
Listening	0.017	0.047
ICT	0.102	0.164
Science	0.194	0.222
Reasoning	0.807	0.808
Reading Speed	0.029	0.059
Precitional Speed	0.193	0.222

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 16, the results only differ marginally between weighted and unweighted indices. Overall, these results — as well as those shown in Table 15 — confirm the statistical significance of the negative effect of radiation exposure on cognitive skills.

³⁷Deviations of the q-values from Equation (5) are due to rounding.

Table 16: Summary index tests

	(1)	(2)
	Unweighted	Weighted
Cognitive skill index	-0.008*** (0.003)	-0.007** (0.003)
Crystallized intelligence index	-0.009*** (0.003)	-0.010*** (0.003)
Fluid intelligence index	-0.006* (0.003)	-0.006* (0.003)
<i>Controls:</i>		
Individual characteristics	Yes	Yes
Municipality characteristics	Yes	Yes
State FE	Yes	Yes

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

E.3 Cluster bootstrap-t procedure (Cameron et al., 2008)

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 17 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 17: Estimates with cluster-bootstrapped standard errors

	(1)	(2)	(3)	(4)
A. Individual test scores				
Math	0.003 (0.004)	0.002 (0.005)	-0.009 (0.009)	-0.011* (0.006)
Reading	-0.001 (0.008)	0.001 (0.015)	-0.008 (0.006)	-0.014* (0.008)
Listening comprehension	-0.003 (0.003)	-0.003 (0.003)	-0.006 (0.004)	-0.009 (0.006)
ICT	0.000 (0.003)	0.001 (0.006)	-0.002 (0.002)	-0.005** (0.003)
Scientific literacy	0.001 (0.001)	0.002 (0.004)	-0.002 (0.002)	-0.004 (0.003)
Reasoning	0.002 (0.004)	0.001 (0.004)	0.001 (0.003)	-0.001 (0.003)
Reading speed	-0.001 (0.007)	0.001 (0.079)	-0.004 (0.007)	-0.008* (0.004)
Perceptual speed	0.003 (0.004)	0.003 (0.004)	-0.002 (0.001)	-0.004** (0.002)
B. Indices				
Cognitive skill index	0.001 (0.008)	0.002 (0.006)	-0.004 (0.004)	-0.008** (0.004)
Fluid intelligence index	0.000 (0.004)	0.002 (0.005)	-0.005 (0.005)	-0.009* (0.005)
Chrystallized intelligence index	0.002 (0.012)	0.003 (0.009)	-0.002 (0.004)	-0.006 (0.005)
<i>Controls:</i>				
Individual characteristics	No	Yes	Yes	Yes
Municipality characteristics	No	No	Yes	Yes
State FE	No	No	No	Yes

Notes: This table corresponds to Columns (5)-(8) in 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 8a shows the distribution of 3.448 measurement points for soil contamination, which is measured by an in-situ gamma ray spectrometer. Due to 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 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 explains clusters of measurement 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 fuer Strahlenschutz, 2016). However, figure 8a reveals 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 measurement 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 8b 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 8a, a uniform distribution is evident across the country. The principal aim of this distribution is the collection of weather data which is representative for the whole country. 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 9, which shows the air concentration of radioactive particles (radionuclides) in four measuring stations in different parts of Germany. The station Brotjacklriegel, a mountain in the south-east, close to the border with the Czech Republic and Austria, is located in the area

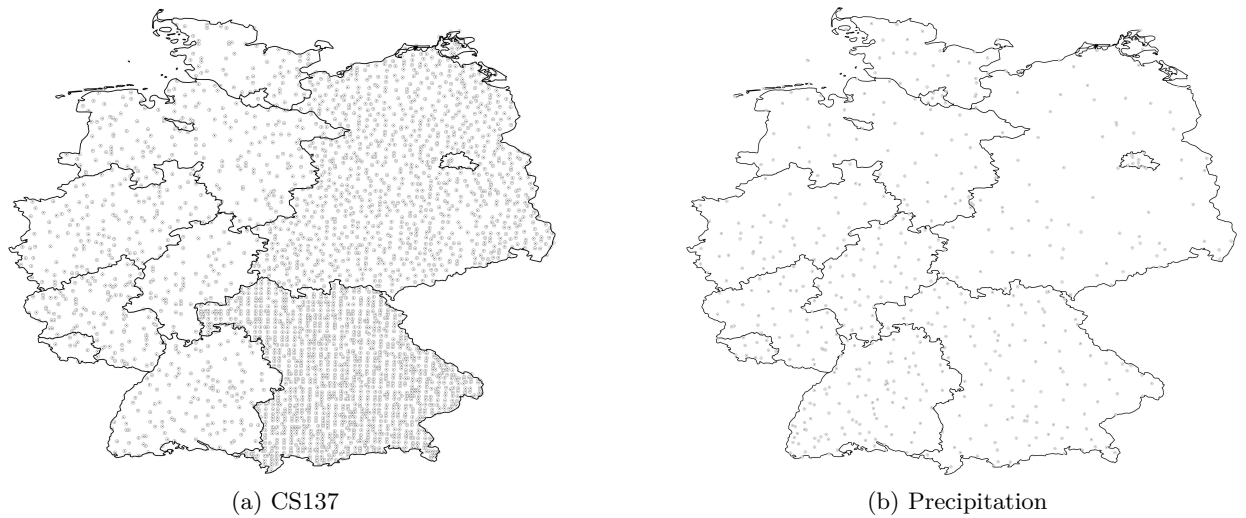


Figure 8: Measurement points. Source: The German Meteorological Service, The Federal Office for Radiation Protection

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 10a. Darker color represents higher precipitation. We determine this period as critical period based on our observations in Figure 9. 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 10b, in contrast, shows the average precipitation between 1981 and 1985. A comparison of Figures 10a and 10b, 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 11a and 11b display the distribution of both variables across space.

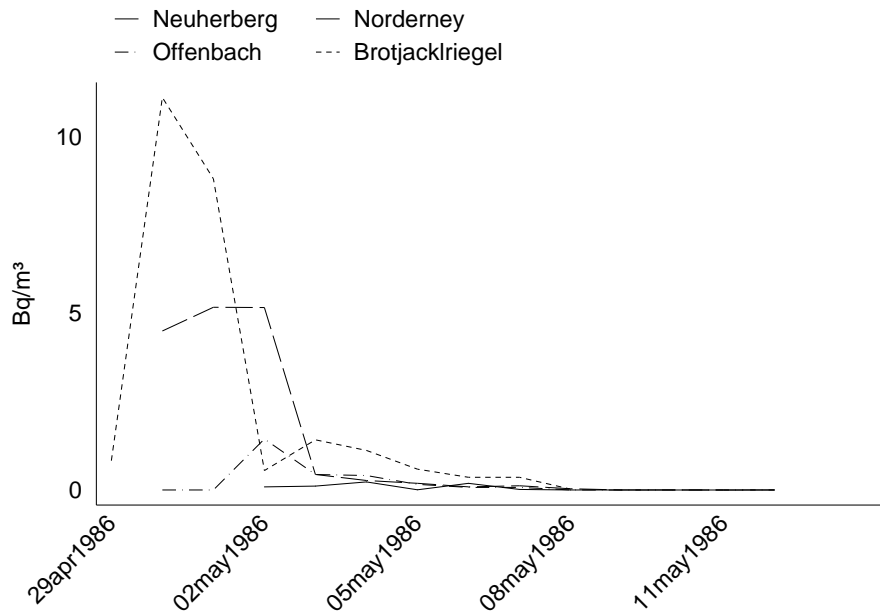


Figure 9: Air concentration of radioactive particles in 1986

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).

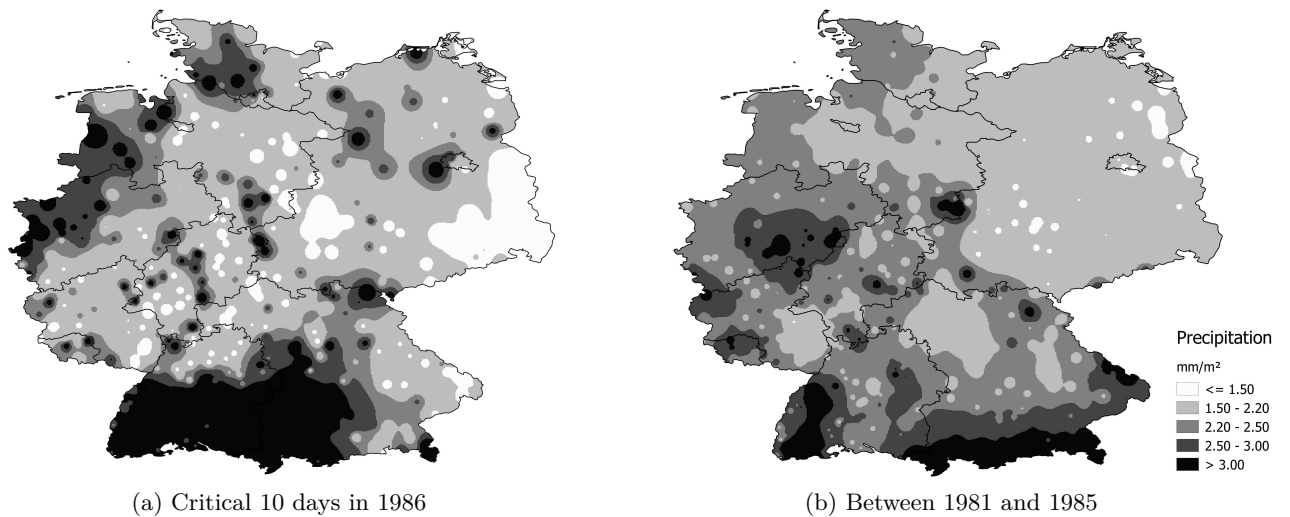
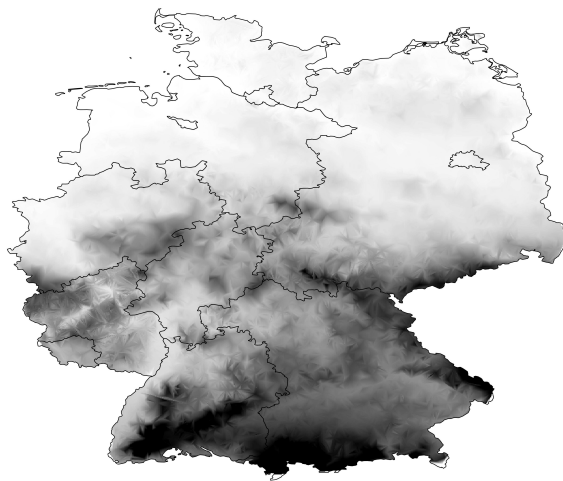
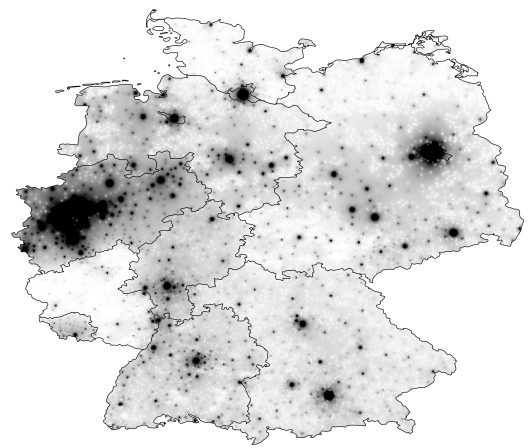


Figure 10: Average daily Precipitation. Source: The German Meteorological Service



(a) Topography



(b) Population density

Figure 11: Altitude and population density, darker means higher. Source: Federal Agency for Cartography and Geodesy

G Control variables

In all specifications that include controls, we use the following variables:

- Personal characteristics

Sex: dummy variable 1) Female 2) Male

Region of origin: dummy variable indicating whether the person was born in 1) FRG or GDR 2) elsewhere

German native speaker: a dummy that equals unity if the person declares that he/she is a German native speaker.

Educational attainment in April 1986: mutually exclusive dummy variables indicating the highest level of education in April 1986: 1) Enrolled but no degree 2) Enrolled, already passed lower secondary and secondary education (Hauptschulabschluss/ Mittlere Reife) 3) Enrolled, already passed upper secondary education (Abitur/ Fachabitur/ Berufsfachschule/ Berufsoberschule) 4) Enrolled, already passed tertiary education (Bachelor/ Master/ Magister/ Diplom/ Staatsexamen/ Promotion) 5) Not enrolled, already passed lower secondary and secondary education 6) Not enrolled, already passed upper secondary education (Abitur/ Fachabitur/ Berufsfachschule/ Berufsoberschule) 7) Not enrolled, already passed tertiary education (Bachelor/ Master/ Magister/ Diplom/ Staatsexamen/ Promotion) 8) Below school age

Economic activity in April 1986: 1) Employed 2) Unemployed

Parental school degree: dummy variables for the highest educational attainment between both parents: 1) Lower secondary (No degree/ Hauptschulabschluss) 2) Secondary (Mittlere Reife/ Fachoberschule) 3) Upper secondary (Abitur/ Fachabitur) (base category: lower secondary)

- Sample design features

Test groups: dummy variables for 1) Math first 2) Reading first 3) Math only 4) Reading only

Year of the competence test: dummy variables for 1) 2010 2) 2012 3) 2014

- Municipality characteristics

Altitude: at the municipality centroid in meters

Minimum altitude: minimum altitude between all municipalities within a county in meters

Average precipitation per year: measured at the municipality centroid, average calculated for 1981 to 1985

Population: log of population living in a municipality in 1997

- State fixed effects:

States: 1) Baden-Wuerttemberg 2) Bavaria 3) Berlin 4) Bremen 5) Hamburg 6) Hesse 7) North Rhine-Westphalia 8) Rhineland-Palatinate 9) Saarland 10) Schleswig-Holstein 11) GDR