Three Empirical Essays in Health Economics

INAUGURALDISSERTATION

zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität Potsdam

vorgelegt von
Florian Wozny, M.Sc.
geboren am 13. Juli 1988 in Soest

-Datum der Disputation: 11.05.2020-

Erstgutachter : Prof. Dr. Marco Caliendo
Zweitgutachter : Prof. Dr. Rainald Borck
Eingereicht im Oktober 2019

Published online in the
Institutional Repository of the University of Potsdam:
https://doi.org/10.25932/publishup-46991
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-469910

Acknowledgments

I am indebted to Marco Caliendo for acting as first supervisor and for providing very valuable feedback and guidance throughout my Ph.D. Moreover, I thank Rainald Borck who immediately agreed to act as the second supervisor. Many thanks go to my co-authors Benjamin Elsner and Nico Pestel for good and enjoyable collaborations. I am also very thankful to all my colleagues at the Institute of Labor Economics (IZA) in Bonn, and specially thankful to my office mates Andreas Lichter, Eric Sommer and Pamela Qendrai for the wonderful working atmosphere. Lydia and Georg, thank you for all the discussions at our kitchen table which aroused my interest in economics. Thank you, Caro, for being my academic role model. Above all, I thank Ann-Ci for her unlimited support.

Contents

List of Figures

List of Tables

General Introduction

Health is a fundamental component of human capital. Healthier workers are physically and cognitively more productive, earn higher wages and are less likely to be absent from work. Population health has a significant positive effect on economic growth (Bloom et al., 2001). Maintaining health is a major policy goal. Health care has become one of the most important economic sectors. On the one hand, this facilitates optimal health care provision. On the other hand, it has become a burden on public budgets as most health care provision is paid by public funds. In order to maintain optimal public health care, effective and cost efficient preventative health care measures need to be combined with the provision of necessary and cost efficient treatments.

Through prevention, there are opportunities to reduce public spending and improve public health. Preventable causes have been estimated to be responsible for 900,000 deaths annually in the United States, or nearly 40 percent of total yearly mortality (Mokdad et al., 2004). Some preventive measures, such as counseling adults to quit smoking or providing influenza vaccination, reduce mortality and are cost efficient (Maciosek et al., 2006). However, general statements about the efficiency of prevention are not possible. Sometimes, preventing illness reduces public spending but in other cases can add to health care costs. Screening costs, for example, can exceed savings from avoided treatments if only a very small fraction of the population would have become ill in the absence of screening (Cohen et al., 2008). Furthermore, screening causes side-effects and thus additional health costs. It is important to analyze the trade-off of costs and benefits of preventive interventions.

Besides prevention, financing of medical treatments offers large potential for cost savings. Reimbursement prices for medication or medical treatments are often negotiated between medical providers and health insurance company foundations or set by public institutions. Although it seems to be obvious that lower prices decreases costs, it might be a misconception if, for example, medical providers reduce supply for unprofitable but necessary treatments and, on the other hand, provide an unnecessary amount of profitable treatments. This could worsen medical supply and at the same time does not result in lower costs.

Causal estimates on how these parameters affect public health outcomes are central inputs to the design of modern health care policies. Obtaining evidence on successful prevention and reimbursement policies is non-trivial, due to a lack of data availability and exogenous variation in the parameters of interest. To overcome the challenge, this dissertation comprises three novel research designs and highlights the interplay of various dimensions in the health care sector. The first two chapters study the demand-side of health care markets by analyzing the impact of pollutants on physical and cognitive health, thus offering potential leeway for preventative health care measures. Focusing on pollutants is of major importance as they create large economic costs through increasing defensive medical spending (Deschênes et al., 2017). I focus on two pollutants that are of major importance in the health care sector. One of them is air pollution, which is the main external source of health problems (WHO, 2018b). The other is radiation. Radiation is of major importance because it is a pollutant mainly emitted in health care by imaging diagnostic, for example (NCRP, 2009). The third chapter studies reallocation processes in hospitals if treatment prices change, thus focusing on the supply-side.

Chapter 1: Radiation and Human Capital after Birth. Chapter one (joined work with Benjamin Elsner) studies the long-term effect of post-natal low dose radiation exposure on cognitive skills. Due to man-made sources, the average person in Europe and America today receives twice the annual dose of radiation compared to the 1980s. Medical treatments like CT scans or x-rays are the main sources of this increase (NCRP, 2009).

The medical literature provides extensive evidence of negative health effects, especially for higher doses of radiation. However, evidence for low dose radiation is less clear as typical sources like medical treatments or working environments are potentially endogenous. The existing economic literature has focused on in-utero exposure, finding significant adverse effects on education and labor market outcomes (Almond et al., 2009; Black et al., 2013). There exists little evidence on the effect of exposure later in life on cognitive skills. Knowing the impact of radiation later in life is important for health policies because a much bigger part of the population is exposed outside the womb. This chapter addresses this gap based on a natural experiment.

Estimations in this chapter are based on regional variation in nuclear fallout after the Chernobyl disaster in 1986, which led to an increase in radiation levels in most of Europe. The chapter relies on the National Educational Panel Study (NEPS), a representative survey of the German population born between 1956 and 1986. NEPS offers three features that are key to our analysis. First, it includes reliable measures of cognitive skills long after school-leaving age, enabling us to analyze the impact of radiation on cognitive skills up until a person's late-fifties. Second, at the time of Chernobyl, the hole sample was already born, allowing us to study the long-term effect of radiation for people who were first exposed later in life. Finally, the survey contains a detailed residential history for each participant, allowing us to link personal information with data on radiation in the respondent's place of residence. We match the NEPS with fine-grained decay-corrected radiation data from a measurement program rolled out by the German government between 1986 and 1989.

The novel combination of the two datasets enables us to run a reduced-form regression of cognitive skills in 2010 and afterwards on the initial level of fallout in a person's place of residence in 1986.

In order to identify a causal effect, we exploit the fact that the degree of soil contamination, caused by Chernobyl, depended on rainfall within a critical ten-day window after the disaster. We show that people who lived in highly-contaminated areas perform significantly worse in standardized cognitive tests 25 years later. This effect is driven by the older cohorts in our sample, whereas we find no effect for people who were first exposed at the ages of 0-7. These results point to significant external costs of man-made sources of radiation.

Chapter 2: Low Emission Zones for Better Health. This chapter (joined work with Nico Pestel) studies health effects of improved air quality by implementing Low Emission Zones. Poor air quality causes about seven million premature deaths each year and health hazards like respiratory and cardiovascular diseases (WHO, 2018a). A major source of ambient air pollution in urban areas is traffic, which is especially harmful as traffic emits pollutants close to humans. However, little is known about potential health benefits from policy interventions aiming at improving air quality in inner cities. In the European Union, a key policy measure to reduce ambient air pollution in inner-cities is the implementation of Low Emission Zones. These are signposted areas where access of vehicles is regulated, typically banning high-emitting vehicles from entering the zone altogether. The main targeted air pollutants are particulate matter and nitrogen dioxide. Both pollutants are linked to a number of respiratory and cardiovascular diseases (Kampa and Castanas, 2008; Schneider et al., 2018).

For identification, we exploit variation in the timing and the spatial distribution of the introduction of new Low Emission Zones across cities in Germany. In a first step, we analyse how Low Emission Zones affected air pollution. In a second step, we show how changes in air quality translate into health outcomes. We use data from the air pollution monitoring system of the German Federal Environment Agency and detailed hospitalization data from the hospital quality reports between 2006 and 2016 which we combine with geo-coded information on the coverage of Low Emission Zones. For every hospital location, we generate catchment areas based on driving time and calculate the share covered by a Low Emission Zone.

Low Emission Zones significantly reduce levels of air pollution in urban areas and improvements in air quality translate into population health benefits. The number of diagnoses related to exposure to air pollution is significantly reduced for hospitals located within or in close proximity to a Low Emission Zone after it becomes effective. The results are mainly driven by reductions in cardiovascular and respiratory diseases. Thus, Low Emission Zones are effective preventive health care measures.

Chapter 3: Reallocation of Hospital Resources. Chapter three studies the reallocation of hospital resources in case treatment prices change at the hospital level. In recent decades, hospital expenses have increased massively (OECD, 2018). At the same time, large-scale reallocation processes of hospital resources occurred. Whether these resource reallocations are caused by healthcare spending has been debated in a so far inconclusive literature (Sloan and Hsieh, 2017). However, understanding the effect of changes in healthcare spending on the reallocation of hospital resources is important because hospital resources are crucial for the quality of healthcare (Needleman et al., 2002).

In this chapter, I show how hospitals reallocated resources between 2006 and 2016 after exogenous treatment price shocks. For my analysis, I use the same dataset for hospitalization as in chapter two and use the same approach to generate catchment areas. I focus on the reallocation between human resources, capital stock as well as changes in services offered to analyze underlying hospital preferences for input and output factors under treatment price shocks. Additionally, I focus on changes of the organizational structure.

In 2005, the introduction of the German Diagnosis Related Groups caused continuous idiosyncratic treatment price shocks at the hospital level until 2010. Starting with hospital individual treatment prices in 2005, prices converged to the federal state average until 2010. Some hospitals were exposed to price increases, while others experienced price reductions. The reform enhanced price decreases and mitigated price increases between 2005 and 2010 for individual hospitals if its number of in-patient cases in 2004 was higher and more severe than in 2003, compared with other hospitals in a federal state.

However, treatment price shocks are endogenous to the allocation of hospital resources. They reflect historic cost structures which might correlate with resource reallocations after the reform, for example. In order to avoid that the result are driven by unobserved heterogeneity, I instrument price shocks by exploiting idiosyncratic changes of weather conditions between 2003 and 2004 in hospital driving time catchment areas. This novel approach is based on unique high-resolution satellite data of population-weighted exposure to the yearly number of days of snow. Days of snow are found in the literature to have a strong positive impact on the number of cost intensive hospital admissions, easily doubling the number of related admissions in winter months (Franklin et al., 1995). Germany is an interesting case for this novel identification strategy due to its changeable weather conditions, which makes it difficult for people to avoid accidents by adjusting their behavior. First-stage results reveal a strong correlation between changes in the number of days of snow and hospital individual treatment prices, driven by snow-related admissions.

The results of the chapter show that higher treatment prices significantly increase the investment in human resources, lead to less privatization and mergers but do not affect the stock of capital. Furthermore, hospitals specialize less and reduce the treatment volume, which indicates supplier-induced demand. Structural differences in pre-treatment characteristics lead to heterogeneous effects. Private and small hospitals are more effected. Effects tend to be persistent even when treatment price shocks vanish. IV estimates reveal that OLS results are biased towards the mean in almost all dimensions, indicating endogenous treatment price shocks.

CHAPTER 1

Radiation and Human Capital after Birth

This chapter studies the long-term effect of post-natal radiation exposure on cognitive skills. We use regional variation in nuclear fallout after the Chernobyl disaster in 1986, which led to an increase in radiation levels in most of Europe. To identify a causal effect, we exploit the fact that the degree of soil contamination depended on rainfall within a critical ten-day window after the disaster. Based on unique geo-coded survey data from Germany, we show that people who lived in highly-contaminated areas in 1986 perform significantly worse in standardized cognitive tests 25 years later. This effect is driven by the older cohorts in our sample, whereas we find no effect for people who were first exposed at ages 0-7. These results suggest that pollution can have adverse effects even when people are first exposed as adults, and point to significant external costs of man-made sources of radiation.[1]

[1] This chapter is based on joint work with Benjamin Elsner. A prior version was published as IZA Discussion Paper No. 11408

1.1 Introduction

The last 40 years have seen a drastic increase in radiation exposure. Today, the average person in Europe and America receives about twice the annual dose of radiation compared with in 1980.[2] This increase is almost entirely due to man-made sources of radiation, namely medical procedures and nuclear power. CT scans, x-rays or mammograms expose patients to low doses of radiation. Moreover, many people were exposed to the nuclear disasters in Chernobyl and Fukushima, whose fallout has been widely distributed. Medical research shows that subclinical radiation damages human cells, which has potential knock-on effects on health and cognition. These effects may occur at all ages. They are particularly strong during critical periods, for example, during gestation or during puberty, although the biological effects are not limited to those periods. The existing literature has largely focused on the effect of in-utero exposure, documenting significant adverse effects of radiation exposure during pregnancy on education and labor market outcomes many years later (Almond et al., 2009; Heiervang et al., 2010; Black et al., 2013). However, there is little evidence on the long-term effects of radiation exposure after birth.

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

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

[2] See NCRP (2009).

and afterwards on the initial level of fallout in a person's place of residence in 1986.

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

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

Our results imply that post-natal exposure to radiation has stronger negative effects than previously thought. Over 25 years, people who lived in an area with a one-standard-deviation higher level of fallout received a cumulative dose equivalent to two mammograms or 40 percent of a CT scan. According to our estimates, this dose reduces cognitive test scores by the equivalent of 0.05-0.08 school years. These effects are similar to those found for cancer patients who underwent radiotherapy (Hall et al., 2004; Pearce et al., 2012), although they are an order of magnitude smaller than the effect of in-utero exposure. For example, Almond et al. (2009) find an effect that is nine times larger. However, it is important to consider the number of people exposed in Germany at different ages. At the time of Chernobyl, around 200,000 fetuses were in the womb during the critical period between 8 and 25 weeks of pregnancy. The birth cohorts in our sample, 1956 to 1985, amounted to 24 million people. Therefore, although the post-natal effects are smaller, they are economically significant because they apply to much larger parts of the population.

This chapter contributes to two strands of literature. First, it fills a gap in the literature on the effect of pollution on human capital. This literature has mainly focused on two effects, namely i) the long-run effect of in-utero exposure on outcomes such as birth weight, crime rates, wages or IQ (Almond and Currie, 2013; Graff Zivin and Neidell, 2013) and ii) the contemporaneous effect of exposure during adulthood on productivity, often measured on the same day.[3] However, there is little evidence on the long-run effect exposure to pollution *after* birth. By showing that constant exposure to radiation after birth has affected cognitive skills over more than two decades, we are among the first papers to document such an effect.

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

The remainder of this chapter is structured as follows. In Section 1.2, we provide the

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

background of the Chernobyl nuclear disaster and the fallout in Germany. In Section 1.3, we summarize the medical literature on the effect of radiation and develop a conceptual framework that guides our analysis. In Section 1.4, we describe the dataset and provide descriptive statistics. Section 1.5 explains the identification strategy and discusses potential threats to identification. In Section 1.6 and 1.7, we present the main results and explore non-linear as well as heterogeneous effects and we carry out extensive robustness checks, before concluding in Section 1.8.

1.2 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.[4] 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 29, 1986. It first entered the country in the south-east and made its way north-west before disappearing over the North Sea on May 8. The fallout comprises four main isotopes, namely caesium-137 (CS137), caesium-134 (CS134), strontium-90 (Sr90) and iodine-131 (I131), which have half-lives of up to 30 years.[5] Among the four isotopes, soil-bounded CS137 is today considered the only relevant source of radiation in Germany that can be ascribed to the Chernobyl disaster (Hachenberger et al., 2017). From 1986 to 1989, the governments of West and East Germany rolled out a comprehensive program to measure radiation across the country. At over 3,000 temporary measuring points, gamma spectrometers measured the radiation of CS137. Based on the decay of the isotopes, all measurements were backdated to May 1986.

The deposition of the fallout varies considerably across regions, and depends on the amount of rainfall within a critical time window. Regions with heavy rainfall while the ra-

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

[5] The half-lives of the four isotopes are eight days (I131), two years (CS134), 28.8 years (Sr90), and 30.2 years (CS137). We will use the abbreviations in parentheses further in the chapter. These do not correspond to the abbreviations used in chemistry, which are ^{137}Cs, ^{134}CS, ^{90}Sr and ^{131}I.

dioactive plume was hanging over Europe received large amounts of fallout whereas regions without rainfall received little to none. Figure 1.1a displays the ground deposition of CS137 in May 1986. Because CS137 rarely occurred in Germany before 1986, the displayed variation is almost entirely due to the Chernobyl fallout. The regions that received the highest level of fallout were Bavaria and Baden-Wuerttemberg in the south as well as parts of the former German Democratic Republic. Across Germany, the level of ground deposition ranges from 0.224 kBq/m^2 to 107 kBq/m^2, whereas soil is officially considered contaminated if the radioactivity exceeds 37 kBq/m^2 (UNSCEAR, 2000). The majority of the population lived in areas with radiation levels below 20 kBq/m^2, although a non-negligible number of people lived in areas with levels much higher than that.[6]

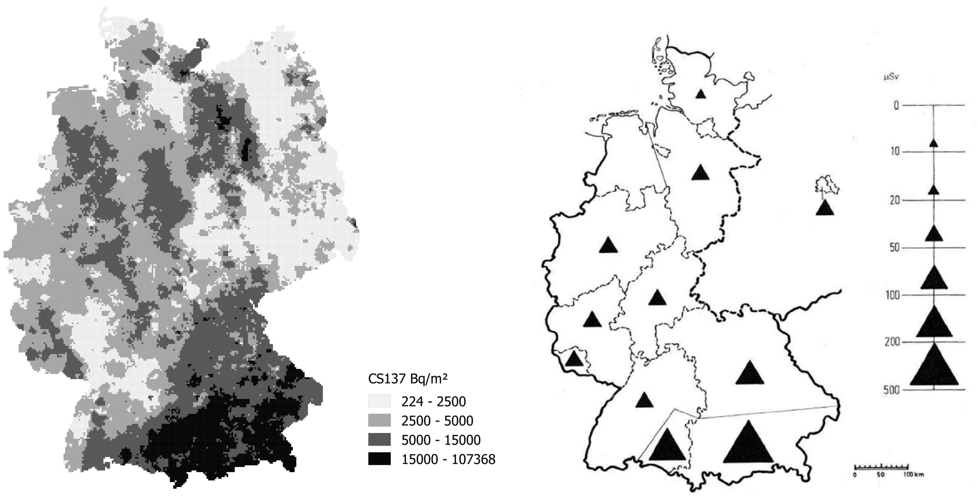
For affected regions, the nuclear fallout represented a quasi-permanent shock to radiation levels. While the air concentration of radioactive particles vanished after a few days, the ground deposition remains in the soil until today. Therefore, a person who has been living for the last 25 years in a highly-affected area has been constantly exposed to a higher dose of radiation than someone living the entire time in a less affected area. In 2010, the first year in which we measure 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 percent of the annual effective dose the average German receives from natural background radiation in one year (2mSv), or the dose received during 30 chest x-rays. However, the effective dose from Chernobyl varied considerably across regions. In Munich, one of the most affected cities, the cumulative effective dose over 25 years was 2.1mSv.[7] Due to the decay of the radioactive matter, the annual effective dose declined over time. The BfS estimates that the dose in the first year, when radioactive particles were in the air and, in general, the radiation was highest, accounted for 21 percent of the cumulative dose over 25 years. In 1987, the dose accounted for 11 percent, and it has been declining at an annual rate of 4 percent since.

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 coun-

[6] See Figure 1.A.1.2 b in Appendix 1.A.1 for the distribution of ground deposition in the German population.

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



(a) Ground deposition of CS137

(b) Information released in 1986

Figure 1.1: Ground contamination in 1986

Notes: These graphs display (a) the ground deposition of CS137 in Bq/m^2 and (b) the information about regional exposure in mSv that was released to the public in 1986. Source: Federal Office for Radiation Protection (Bundesamt für Strahlenschutz), German-Swiss Association for Radiation Protection (Fachverband für Strahlenschutz e.V.)

try, 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 II, which reported about high levels of radioactive matter being emitted from an exploded nuclear power plant in Ukraine. In the same newscast, the Federal Minister of the Interior, Friedrich Zimmermann, stated that, due to the distance to Ukraine, there was no danger for the German population. However, two days later, after high radiation levels were measured in several parts of the country, the government of the Federal Republic of Germany (FRG) introduced radiation limits on foods and warned the population of the consumption of dairy produce, vegetables, mushrooms and game, which were potentially contaminated. In the following days, contaminated food was discarded and public swimming pools and playgrounds were temporarily closed. Despite these measures, the German government maintained its official communication that the increased radiation did not present a health hazard to the population. The information policy differed considerably between the FRG and the German Democratic Republic (GDR). In the GDR, no comparable measures were put in place. Quite the opposite, after the accident and the collapse of demand in the FRG, agricultural products intended for export to the FRG were supplied to the market in the GDR.

While the German population was generally informed about the radioactive fallout, they had little knowledge about the levels of fallout in particular areas. Figure 1.1b shows a map released by the German-Swiss Association for Radiation Protection in 1986, which displays the average exposure in mSv in twelve large regions. A detailed map, such as the one shown in Figure 1.1a 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.

1.3 Radiation and cognitive test scores: conceptual framework

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

1.3.1 Exposure to radiation

Humans can be exposed to radiation in three ways, namely through inhaling radioactive particles, ingesting contaminated foods, as well as external exposure, whereby radiation affects the body if a person is present in a place with a given level of radioactivity in the environment. Exposure to radiation through air and ground can be directly assigned to, and therefore strongly correlated with, a person's place of residence (Clark and Smith, 1988). By contrast, exposure through food may not necessarily result from contamination in the same locality, given that the food might have been produced elsewhere.

In the northern hemisphere, the average yearly exposure to natural radiation is 2.4 mSv, of which 52 percent is through inhalation, 12 percent through ingestion, and 36 percent through terrestrial and cosmic radiation (UNSCEAR, 2008). The degree of exposure differs between people and depends on their daily activities and diet. For example, people who spend more time outdoors are more exposed to cosmic radiation than those who spend most of their time indoors, while people who are physically active, and therefore breathe more, have a higher exposure through inhalation.

1.3.2 Impact on the human body

Radiation affects the human body through ionization, a process that damages the DNA and can lead to the dysfunction or death of cells. When it collides and reacts with the DNA in a cell, radiation can directly or indirectly damage the DNA. Radiobiology theory posits that a marginal increase in radioactivity linearly increases the probability that a cell is hit by an electron. A linear relationship emerges because during ionization the release of electrons follows a random process, whereby each cell has an equal likelihood of being hit. A marginal increase in radioactivity increases this likelihood and leads to a greater number of cells being hit (Brenner et al., 2003). The human organism has the capacity to repair damaged DNA. However, if the DNA is not fully repaired, the cell may continue to regenerate and differentiate, thereby passing on the damaged DNA to future cell generations. This process can lead to mutations as well as the dysfunction of cells. The greater the number of affected cells and the longer the observation period, the more likely that a critical mass of dysfunctional cells affects the functioning of organs and therefore leads to adverse health effects.

Impact on health. The medical literature provides ample evidence of negative health effects. These effects can be either *deterministic*, whereby exposure to radiation almost inevitably affects a person's health, or *stochastic*, in which case radiation affects the likelihood of developing a health condition. Deterministic effects only result from high doses of radiation such as those encountered by survivors of the Hiroshima nuclear bomb or soldiers who cleaned up the nuclear waste in Chernobyl. By contrast, a low dose of radiation, defined as a short-term dose below 100 mSv, only induces stochastic health effects. At such levels, an increase in the dose raises the probability that a person experiences health problems later in life, but does not lead to the immediate dysfunction of organs (OECD, 2016). The medical literature provides evidence of the existence of stochastic health effects such as heart disease, stroke, digestive diseases, and respiratory diseases (Preston et al., 2003). People in at-risk occupations, for example, workers at nuclear power plants, who receive an additional dose between 1 mSv and 2.5 mSv per year, are shown to have a higher cancer risk (UNSCEAR, 2008).

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

Although the impact of radiation on cognition is strongest during pregnancy, it can unfold throughout a person's life. Recent research shows that radiation affects cell regeneration in the hippocampus, the part of the brain that governs several types of memory, in particular crystallized intelligence and learning (Squire, 2009; Supekar et al., 2013). Several studies show that exposure to radiation slows down the regeneration of brain cells, which in turn impairs cognitive performance. People who have been exposed to low-dose radiation during medical treatments are more likely to suffer from cognitive impairments several months to years later (Hall et al., 2004; Douw et al., 2009; Monje and Dietrich, 2012) and have a higher risk of developing dementia (Greene-Schloesser and Robbins, 2012). Therefore, it is biologically plausible that radiation affects cognition even if the exposure occurs long after birth.[8]

1.3.3 Conceptual framework and predictions

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

$$y = F[I(R,B), H(R,B), B]. \tag{1.1}$$

A cognitive test score y is produced with three inputs: a person's intelligence I, a person's health H, as well as any choices that people make in response to being exposed, summarized by $B = B(R)$. We think of B as compensating behaviors aimed at limiting or counteracting the impact of radiation. There are many possible behavioral responses; for example, investment 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 (1.1) yields the proportional change of a test score in response to a change in levels of radiation,

$$\frac{dy}{dR} = \underbrace{\frac{\partial F}{\partial I}\frac{\partial I}{\partial R} + \frac{\partial F}{\partial H}\frac{\partial H}{\partial R}}_{\substack{\text{direct effects} \\ \text{(cognition, health)}}} + \underbrace{\frac{\partial F}{\partial B}\frac{\partial B}{\partial R} + \frac{\partial F}{\partial I}\frac{\partial I}{\partial B}\frac{\partial B}{\partial R} + \frac{\partial F}{\partial H}\frac{\partial H}{\partial B}\frac{\partial B}{\partial R}}_{\text{behavioral responses}} . \tag{1.2}$$

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.

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

Equation (1.2) allows us to generate hypotheses about the sign of each channel, although the sign of the overall effect remains ambiguous. In light of the existing evidence, the direct effects are either negative or zero. Low-dose radiation may have a negative impact on health and cognition unless the dose is too small to have any impact at all. On the other hand, the sign of the behavioral responses is either positive or zero. There is plenty of anecdotal evidence of behavioral responses to the Chernobyl disaster. According to a German survey from 1987, many people initially followed the government's recommendations to avoid certain foods and keep their children inside during the weeks after the radioactive rainfall (Peters et al., 1987). In addition, a study from Austria by Halla and Zweimüller (2014) shows that families responded to the fallout by having fewer children and reducing mothers' labor supply. However, as shown by Renn (1990), many behavioral responses, especially changes in diet and exercise habits, were fairly short-lived.

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

Equation (1.2) also helps to interpret the estimates. Our regression, assuming that radiation exposure is exogenous, allows us to identify the *total* effect of radiation exposure on test scores, which comprises direct effects as well as compensating behaviors. If one was interested in the importance of a particular channel, this would require controlling for all other channels or finding a quasi-experimental design in which the remaining channels are absent. In our analysis, while we cannot fully disentangle the direct and indirect channels, our data allow us to test whether some plausible behavioral channel have an influence by testing whether $\frac{\partial B}{\partial R} = 0$.

1.4 Data and Descriptive Statistics

We link rich individual-level survey data with geo-coded information on radiation in a person's municipality of residence in May 1986. In this section, we describe the construction of the dataset as well as the measurement of cognitive skills, and present descriptive statistics. We limit the description of the dataset to the most important aspects. In addition, in Appendix 1.A.1, 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.

1.4.1 The NEPS data

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

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

The NEPS SC6 includes all respondents of ALWA who were willing to enter the panel and be surveyed every year (N=8,997). Among those who agreed to be included, 6,572 actually participated.[9] A comparison of the ALWA subsample with the German Microcensus shows that 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.

1.4.2 Estimation sample

Our sample includes all survey participants who were born *before* Chernobyl. We exclude participants born after Chernobyl because the survey only sampled birth cohorts up to December 1986, leaving us with few participants who were born after Chernobyl. Moreover, because we are interested in the effect of post-natal exposure, excluding them ensures

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

that our estimates are not confounded by exposure in utero, which operates through a different biological channel. Overall, we can link the municipality of residence in May 1986 for 5,844 participants. For the remaining 728 participants, we could not link the data due to missing municipality keys (402 obs.) or because they lived abroad in May 1986 (326 obs.). Observations with missing municipality keys include 140 participants born after April 1986.

To reduce classification error, we drop respondents who moved in May 1986 (34 obs.), for whom we cannot determine whether they moved before or after the radioactive plume reached Germany. We also drop all respondents who did not participate in the competence tests (1,265 obs.), as well as all participants for whom information on personal characteristics is missing (105 obs.). Our final estimation sample comprises 4,440 observations. In Appendix 1.A.1, we provide a detailed description of the sample design and the actions taken by the interviewers to minimize recall error when eliciting the residential history. Moreover, in order to address concerns about the representativeness of the estimation sample, we perform a series of balancing tests in Appendix 1.A.2, which suggest that the missing information is unsystematic.

1.4.3 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 1.A.1 provides a detailed description of each test. In the empirical analysis, we use each test score as a separate outcome. In order to make the estimates comparable across outcomes, we standardize the test scores to a mean of zero and a standard deviation of one. Moreover, given that the test scores measure different aspects of the latent variable cognitive skills, we construct a standardized cognitive skills index that allows us to estimate the overall effect of radiation on the latent factor cognitive skills. To construct the index, we first sum over all eight standardized test scores, and then standardize this sum to a mean of zero and a standard deviation of one. Using the same standardization, we construct sub-indices for skills governed by crystallized intelligence (math, reading, science, listening, ICT) and fluid intelligence (reading speed, perceptual speed, reasoning).

1.4.4 Municipality- and County-level data

Data on ground deposition. Our regressor of interest is the ground deposition of CS137 in kBq/m^2 in May 1986, which we use as proxy for Chernobyl-induced radiation in Germany. The regional concentration of CS137 is strongly correlated with other Chernobyl-induced sources of radiation such as I131 or Sr90 (Hou et al., 2003), although CS137 is easier to measure and, due to its long half-life, mainly responsible for the long-run exposure of the population (International Atomic Energy Agency, 2006).

The Federal Office for Radiation Protection (Bundesamt für Strahlenschutz, BfS) provided us with geo-coded data for the soil surface contamination in Germany at 3,474 measurement points in May 1986. The data were compiled by the BfS following a comprehensive measurement program rolled out between 1986 and 1989. Measurements taken after May 1986 were backdated based on the decay of CS137. For each municipality centroid, we calculate the radiation level as the inverse-distance weighted average from the four closest measuring points. After 1989, no comparable radiation data are available. Therefore, we know the *initial* level of radiation in area, but we have no information how radiation levels developed between 1989 and 2010. It is possible to calculate the approximate radiation level based on the decay of CS137, although to determine the exact level we would need to know the extent to which the radioactive matter was washed into deeper layers of soil.

Linkage between individual and regional data. We link the radiation data for 1986 with the individual survey data based on the respondents' municipality of residence in May 1986, using the radiation level in the centroid of the municipality. This linkage provides us with a measure of potential exposure to the post-Chernobyl radiation for each person in the sample.[10] 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 1.A.2 we perform robustness checks, which show that the results are not driven by the linkage procedure.

Additional data. We supplement our dataset with municipality- and county-level data on geographic conditions and population characteristics. We obtained data on precipitation, altitude and population size at the municipality level and data on minimum altitude and the composition of the population at the county level. In addition, we obtained data on dementia incidence from the German Hospital Quality Reports provided by Destatis. This dataset includes all dementia cases that were diagnosed in hospitals in a given municipality between 2006 and 2016. In Appendix 1.A.1, we provide a detailed description of all variables used in this chapter.

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

1.4.5 Descriptive statistics

Table 1.1 displays the descriptive statistics of the main variables used in the regression. In 1986, the average person in the sample was 19 years old, with ages ranging from zero to 30 years. 36 percent of the sample, predominantly the older cohorts, were employed at the time, while another 43 percent were enrolled in education, and 1 percent were unemployed. The share of people who lived in the GDR represents 18 percent 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 percent of the sample were no longer in education in April 1986: 4 percent had a lower secondary or secondary, while 28 percent and 13 percent, respectively, had an upper secondary or tertiary degree. On the other hand, 43 percent were still in education, most of whom had not yet finished a degree (31 percent of the sample). 10 percent of the sample were enrolled in 1986 but had already passed lower secondary or secondary education, while 1 percent had passed upper secondary education.

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

The fourth set of statistics describe the cognitive test scores. Two features are noteworthy here. First, each test has a different metric, resulting in differences in means and standard deviations. Without a standardization, the estimates will be difficult to interpret and compare. Second, the number of observations differs between tests, which is due to design features of the NEPS (see Section 1.4.3 and Appendix 1.A.1).

Panel B displays the municipality-level characteristics. With the exception of dementia incidence, the statistics were computed across individual observations in the estimation sample. The mean ground deposition of CS137 in May 1986 amounts to 5.18 kBq/m^2. The standard deviation, which is larger than the mean, points to a significant variation in ground deposition across Germany.[11]

The level of precipitation represents the average rainfall in May in the five years preceding the Chernobyl disaster, i.e. 1981-1985. In 1986, the average person lived in a medium-sized municipality with 282,000 inhabitants, although municipality sizes vary between 5,000 and over 3 million. The last row of Panel B displays the number of dementia cases between

[11] See Appendix 1.A.1 for an illustration of the distribution of the ground deposition across municipalities.

2006 and 2016 that were diagnosed in hospitals. Across municipalities, this number varies from zero to over 6,000, with a mean of 148.

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

Table 1.1: Descriptive statistics of the main variables

	Mean	SD	min	max	N
A. Individual-level data					
Personal characteristics					
Age in 1986	19.05	8.20	0.00	30.43	4440
Female	0.51	0.50	0.00	1.00	4440
Native speaker	0.98	0.15	0.00	1.00	4440
GDR	0.18	0.39	0.00	1.00	4440
Unemployed in April 1986	0.01	0.12	0.00	1.00	4440
Employed in April 1986	0.36	0.48	0.00	1.00	4440
Educational attainment in April 1986					
Not of school age yet (less than 7 years old)	0.12	0.33	0.00	1.00	4440
No degree, lower secondary, secondary	0.04	0.19	0.00	1.00	4440
Upper secondary	0.28	0.45	0.00	1.00	4440
Tertiary	0.13	0.33	0.00	1.00	4440
In school or college education,	0.43	0.49	0.00	1.00	4440
no degree	0.33	0.47	0.00	1.00	4440
already attained lower secondary, secondary	0.10	0.31	0.00	1.00	4440
already attained upper secondary	0.01	0.09	0.00	1.00	4440
Highest parental education					
Lower secondary	0.52	0.50	0.00	1.00	4440
Secondary	0.27	0.44	0.00	1.00	4440
Upper secondary	0.21	0.41	0.00	1.00	4440
Test Scores					
Math	11.32	4.75	0.00	21.00	2652
Reading	27.06	7.45	0.00	39.00	2666
Reading Speed	38.19	8.34	0.00	51.00	3611
Scientific literacy	19.00	5.29	0.00	30.00	3286
ICT	41.20	13.62	0.00	66.00	3312
Reasoning	8.94	2.38	0.00	12.00	3169
Listening comprehansion	75.82	7.97	0.00	89.00	3172
Perceptual Speed	34.68	8.07	0.00	82.00	3170
B. Municipality-level data					
Caesium137 kBq/m^2 (01. May 1986)	5.18	5.87	0.50	62.10	4440
Average Caesium137 kBq/m^2 (until 2010, decay corrected)	3.89	4.41	0.38	46.64	4440
Precipitation mm/m^2 (yearly average, 1981-1985)	3.09	0.84	1.30	8.00	4440
Altitude in meter	201.59	176.69	0.00	850.00	4440
Minimum altitude in meter in county	138.73	139.78	-1.00	660.00	4440
Population/1000	281.67	676.43	5.00	3420.00	4440
B. County-level data					
Tertiary degree/Population	0.07	0.03	0.03	0.16	4440
Working population/Population	0.44	0.07	0.33	0.61	4440
18-65 years old/Population	0.65	0.03	0.59	0.71	4440

Notes: This table displays the descriptive statistics for the variables used in the analysis. The number of observations varies between tests due to the survey design. See Appendix 1.A.1 for a comprehensive description of the testing procedure. The data underlying the statistics in Panel B are measured at the municipality level and in C at the county-level, although the statistics themselves are computed at the individual level. The statistics for dementia cases are measured at the municipality-level.

1.5 Empirical Strategy

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

1.5.1 Empirical model

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

$$y_{ims} = \alpha + \beta\, CS137^{86}_{ms} + \boldsymbol{X'_{im}}\gamma + \delta_s + \varepsilon_{ims}. \tag{1.3}$$

The cognitive test score y_{ims}, measured in and after 2010, of person i who resided in municipality m in state s in May 1986 is regressed on the ground deposition of CS137 in the same municipality in May 1986. The vector $\boldsymbol{X_{im}}$ controls for pre-treatment characteristics of individuals and municipalities as well as design features of the survey. At the individual level, it includes controls for gender, a quadratic in age, a dummy for whether a person is a German native speaker, an indicator whether the person was born in Germany, parental education, education in 1986, and employment status in 1986. It also includes municipality and county characteristics, namely the average daily rainfall between 1981 and 1985, altitude, log population, as well as the share of the population in a county with a tertiary education, the share of people working and the share of people aged 18-65.[12] 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.[13] In some specifications, we also control for state fixed effects, δ_s.[14] The error term ε_{ims} summarizes all determinants of cognitive test scores not captured by the regressors.

[12] The controls for altitude include two variables, namely the altitude at the municipality centroid as well as the minimum altitude in a given county. The combination of these two variables has been shown to be a determinant of orographic rainfall (Houze, 2012), which in turn has been shown to increase the level of fallout (Yasunari et al., 2011). Appendix 1.A.1 provides further details on the control variables.

[13] See Appendix 1.A.1 for a description of the survey design.

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

In line with the conceptual framework in section 1.3, the coefficient β measures the *total* effect of the radiation level in 1986 on cognitive test scores. A higher initial radiation in a municipality in 1986 leads to a higher average radiation between 1986 and 2010, which in turn leads to a higher potential exposure to radiation. In addition, β contains the direct biological channels as well as behavioral responses to the fallout, such as changes in diet, exercise habits or internal migration. While we do not observe whether a higher potential exposure leads to a higher actual exposure, an estimate different from zero $\hat{\beta} \neq 0$ provides indirect evidence that it does.

Given the nature of our data and estimation strategy, we need to address several challenges to statistical inference. To account for potential cross-sectional dependence of the error terms, we cluster the standard errors at the county level. We also test for spatial autocorrelation in the cognitive skills index. Based on Moran's I, we fail to reject the null hypothesis of zero spatial autocorrelation for both variables.[15] In Appendix 1.A.5, we undertake several steps to assess the robustness of our inference. We perform permutation tests, and allow for clustering at the state level by performing a bootstrap-t-procedure (Cameron et al., 2008). We also account for potential multiple hypothesis testing with a summary index test (O'Brien, 1984; Anderson, 2008) and a step-down adjustment of standard errors (Benjamini and Hochberg, 1995).

1.5.2 Identification

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

This identification problem is prevalent in studies concerning the impact of pollution. Most studies on *in-utero* exposure overcome this problem by exploiting critical periods during which pollution affects a fetus. These studies typically compare people exposed to the pollutant during a critical period to people in the same locality who were also exposed, although not during the critical period. The confounding effect of residential sorting is differenced out with location fixed effects. However, for the effect that we want to uncover, the effect of constantly higher exposure over 25 years, this identification strategy is not suitable

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

because there are not many pronounced critical periods during adolescence and adulthood. Instead, our identification needs to rely on cross-sectional variation. The estimating Equation (1.3) compares the cognitive skills of people with similar observables who were differentially exposed because they lived in different places at the time of the disaster.

To estimate a causal effect, we pursue two identification strategies. First, we control for many potential determinants of residential sorting and show that conditional on these controls the sample is balanced on observable characteristics. Second, we apply an instrumental variable approach that exploits idiosyncratic rainfall in a critical time window after the disaster.

1.5.2.1 Selection on Observables

To corroborate the identifying assumption, in Table 1.2 we perform balancing tests whereby we compare pre-determined characteristics of people who lived in areas with above- and below-median levels of fallout in May 1986. To be conservative, we report conventional standard errors, which make discovering a significant difference more likely. The comparison of the raw data in Columns (1)-(3) indeed points to residential sorting. People with a low education as well as those with less-educated parents were more likely to live in areas that received a higher level of fallout. Panel B provides some potential reasons for this sorting pattern. Municipalities with above-median levels of fallout tend to have a higher altitude, are less populated, and have more rainfall. In other words, less-skilled Germans tend to live in more rural areas, and rural areas received a greater level of nuclear fallout after the Chernobyl disaster due to their altitude and rainfall levels.

In Columns (4)-(9), we test whether the sample is balanced conditional on controls. As shown in Columns (4)-(6), controlling for altitude, rainfall and population size cannot fully eliminate residential sorting. In Columns (7)-(9), we additionally control for state fixed effects, restricting the identifying variation to within the state of residence in May 1986. Conditional on these controls, the sample is balanced on all observable characteristics, which is why our preferred specification will include controls for municipality-level characteristics as well as state fixed effects. The state fixed effects also address an important institutional feature in the German education system, namely the fact that states have a high degree of autonomy in their education policy. This manifests itself in significant differences in PISA test scores between states. We account for these differences by including state fixed effects.

1.5.2.2 Instrumental Variable Strategy

While the sample is balanced on a large number of observables, our estimate of β may be confounded by unobserved differences between individuals. To further strengthen our causal identification, we apply an instrumental variable approach. Following a commonly-used method in physics (Pálsson et al., 2006), we predict the amount of fallout based on the interaction of two forces, namely the amount of rainfall while the plume was hanging above a place ($rain_m$) and the amount of radioactive matter in the plume in a given place ($matter_m$).[16] It is intuitive why both forces determine the level of fallout. The only way that an area can receive fallout is through rainfall: for a given amount of available radioactive matter in the plume, areas with higher rainfall receive more fallout. However, because the matter is rained down, the amount of radioactive matter changes as the plume moves. Given that large amounts of radioactive matter were rained down in the south of Germany, less matter could be rained down in North Germany.

The first stage. In the first stage, we predict the level of CS137 based on the instrument, the log of the interaction of rainfall and radioactive matter as well as all the controls included in Equation (1.3),

$$CS137^{86}_{ms} = \lambda_0 + \lambda_1 \ln(rain_m \times matter_m) + \boldsymbol{X'_{im}}\boldsymbol{\kappa} + \rho_s + \eta_{ims}. \tag{1.4}$$

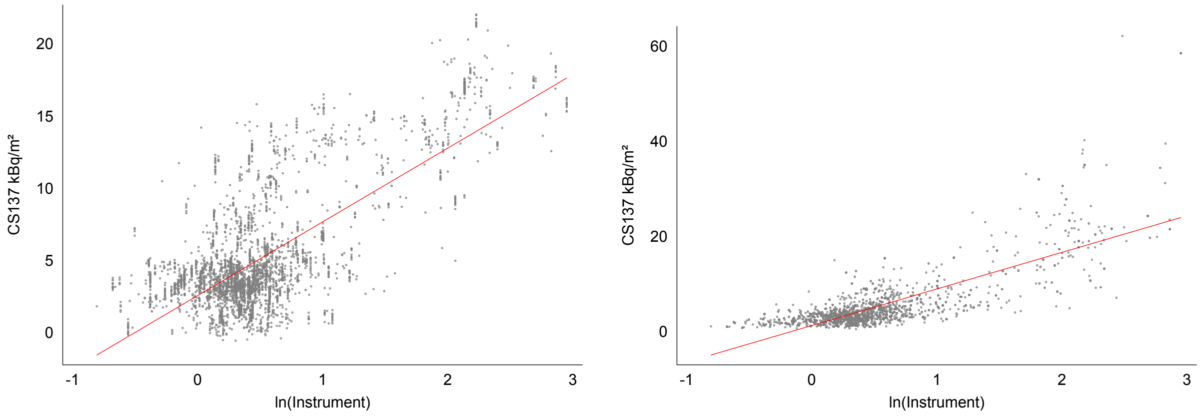
We measure $rain_m$ as the cumulative rainfall in a narrow time window between April 29 and May 8 in 1986 and interact this measure with the amount of radioactive matter in the plume. Importantly, in Equation (1.4) we control for the average level of rainfall between April 29 and May 8 in the years from 1981 to 1985.[17] Therefore, our instrument predicts the level of fallout based on *abnormal* rainfall levels, that is, the deviation in rainfall from the 5-year average within the critical ten days in early May.

Figure 1.2 displays the raw correlation between the instrument and the level of fallout (Panel a) and the correlation after controlling for indivdual, municipality and county characteristics and state fixed effects (Panel b). Even after adding controls, the correlation is strong and has the expected positive sign. In Table 1.A.3.1 in Appendix 1.A.3, we report the first-stage coefficients and F-statistics for all outcomes and multiple specifications. In the first-stage regressions with all controls, the F-statistic of the excluded instrument ranges between 31 and 56, which rules out a weak instrument problem.

Exclusion restriction. In order to deliver a causal estimate, the instrument has to satisfy the exclusion restriction that the rainfall on 10 days in early May 1986, conditional on the *average* level of rainfall in early May and other controls, affects cognitive test scores only

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

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



(a) Raw correlation

(b) Correlation with controls

Figure 1.2: First-stage correlation

Notes: Panel (a) displays the first-stage correlation between the instrument $\ln(rain_m \times matter_m)$ *and the amount of fallout CS137. In Panel (b) we control for the indivdual, municipality and county characteristics as well as state fixed effects mentioned in section 1.5.1. The graph plots the residuals with the means of both axes added in.*

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

To corroborate the exclusion restriction, we perform balancing and falsification tests. In the balancing tests, we regress pre-treatment characteristics on the instrument. A significant coefficient would indicate that the instrument is correlated with the error term in Equation (1.3) and, thus, invalid. The results, displayed in Table 1.A.3.2 in Appendix 1.A.3, show that once we control for state fixed effects, the instrument does not predict pre-treatment characteristics. This finding is consistent with, although not a proof for, the instrument being as good as randomly assigned.

We further perform falsification tests based on the reduced form, i.e. a regression of the outcome on the instrument. Rather than using rainfall between April 29 and May 8, 1986, we compute the instrument based on rainfall on the same days in 1987 and 1988.[18] This is a diagnostic test for the validity of the instrument. If we found significant effects after

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

1986, this would indicate that our instrument picks up a rainfall pattern that is spuriously correlated with the outcome. This would be evidence against the exclusion restriction.

Table 1.A.3.3 in Appendix 1.A.3 shows that while the reduced-form coefficients for 1986 are negative, large, and highly significant, they are considerably smaller, statistically insignificant and in most cases positive when the instrument is based on rainfall in later years. Among the 22 reduced-form regressions we perform, only one coefficient is statistically significant at the 5 percent-level, which is consistent with random sampling variation around a true effect of zero.

Further threats to identification. Besides residential sorting and unobserved heterogeneity, there are at least three additional challenges to identification. One challenge is selective attrition. Radiation can increase the risk of dying from cancer, potentially resulting in a selected estimation sample. Likewise, not all respondents completed all cognitive skills tests, and this non-participation is potentially systematic. Finally, the linkage of radiation data with individual-level survey data introduces measurement error, because we only know the potential exposure in the person's municipality of residence, but neither the ground deposition in the exact location of residence nor the person's actual exposure. We address these challenges through as a series of robustness checks. We discuss the implications of these tests along with the main estimation results in the next section.

Table 1.2: Balancing tests based on pre-determined characteristics

	Raw data			Control municipality charac.			State FE, Municipality charac.		
	Mean		Diff (2)-(1)	Mean		Diff (5)-(4)	Mean		Diff (8)-(7)
	Below	Above		Below	Above				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Individual characteristics									
Age in 1986	18.947	19.137	0.190	-0.290	0.273	0.563**	-0.189	0.177	0.366
	(0.178)	(0.170)	(0.246)	(0.178)	(0.168)	(0.245)	(0.178)	(0.168)	(0.245)
Female	0.517	0.499	-0.018	0.002	-0.002	-0.004	0.002	-0.002	-0.004
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
Native speaker	0.974	0.980	0.006	-0.002	0.001	0.003	-0.000	0.000	0.000
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Employed in April 1986	0.363	0.351	-0.012	-0.007	0.007	0.014	-0.003	0.003	0.006
	(0.010)	(0.010)	(0.014)	(0.010)	(0.010)	(0.014)	(0.010)	(0.010)	(0.014)
Unemployed in April 1986	0.014	0.013	-0.002	0.002	-0.002	-0.003	0.002	-0.002	-0.004
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
If employed: Qualified or highly qualified job before May 1986	0.519	0.524	0.005	-0.006	0.006	0.013	-0.005	0.004	0.009
	(0.017)	(0.017)	(0.025)	(0.017)	(0.017)	(0.024)	(0.017)	(0.017)	(0.024)
Children before 1986	0.178	0.154	-0.024**	-0.008	0.008	0.016	-0.005	0.005	0.011
	(0.008)	(0.008)	(0.011)	(0.008)	(0.007)	(0.011)	(0.008)	(0.007)	(0.011)
Older siblings	0.521	0.546	0.024	-0.000	0.000	0.000	-0.005	0.005	0.010
	(0.011)	(0.011)	(0.016)	(0.011)	(0.011)	(0.015)	(0.011)	(0.011)	(0.015)
Educational attainment in April 1986									
Lower secondary and secondary	0.032	0.043	0.011**	-0.002	0.002	0.003	-0.001	0.001	0.002
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
Upper secondary	0.280	0.289	0.009	-0.004	0.004	0.007	0.001	-0.001	-0.002
	(0.010)	(0.009)	(0.014)	(0.010)	(0.009)	(0.014)	(0.010)	(0.009)	(0.013)
Tertiary	0.137	0.116	-0.021**	-0.002	0.002	0.004	-0.002	0.002	0.004
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
In school or college education	0.424	0.434	0.009	0.001	-0.001	-0.002	-0.002	0.002	0.003
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
In education, already attained lower secondary and secondary	0.103	0.104	0.001	-0.008	0.008	0.016*	-0.008	0.008	0.016*
	(0.007)	(0.006)	(0.009)	(0.006)	(0.006)	(0.009)	(0.006)	(0.006)	(0.009)
In education, already attained upper secondary	0.010	0.006	-0.005*	0.001	-0.001	-0.003	0.001	-0.001	-0.002
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
In education, already attained tertiary	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Not of school age yet (less than 7 years old)	0.127	0.119	-0.008	0.006	-0.006	-0.012	0.004	-0.004	-0.007
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
Highest parental education									
Lower secondary education	0.467	0.575	0.108***	-0.006	0.005	0.011	-0.006	0.005	0.011
	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)	(0.011)	(0.010)	(0.015)
Secondary education	0.288	0.249	-0.039***	-0.004	0.003	0.007	-0.002	0.002	0.004
	(0.010)	(0.009)	(0.013)	(0.010)	(0.009)	(0.013)	(0.010)	(0.009)	(0.013)
Upper secondary	0.246	0.176	-0.069***	0.009	-0.009	-0.018	0.008	-0.007	-0.015
	(0.009)	(0.008)	(0.012)	(0.009)	(0.008)	(0.012)	(0.009)	(0.008)	(0.012)
B. Municipality characteristics									
Caesium137 kBq/m^2 (01. May 1986)	2.283	7.905	5.622***						
	(0.014)	(0.150)	(0.150)						
Altitude in meter	141.419	258.079	116.660***						
	(2.852)	(4.048)	(4.952)						
Minimum altitude in meter in county	88.299	186.072	97.773***						
	(1.705)	(3.459)	(3.856)						
Population	438.286	134.627	-303.659***						
	(19.409)	(5.994)	(20.313)						
Precipitation in mm/m^2	2.911	3.264	0.353***						
	(0.021)	(0.014)	(0.025)						
GDR	0.289	0.082	-0.207***						
	(0.010)	(0.006)	(0.011)						
N	2150	2290							

Notes: This table displays the pre-treatment characteristics of individuals (Panel A) and municipalities (Panel B) in areas with above- and below-median ground deposition of CS137. Columns (1) and (2) display the raw means and standard deviations, whereas Columns (4) and (5) as well as (7) and (8) display the residual means and standard deviations after conditioning on municipality characteristics and state fixed effects. In Columns (3), (6), and (9), we perform t-tests for equality in means. Conventional standard errors of the test statistics are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

1.6 Radiation and Cognitive Skills: Results

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

1.6.1 OLS Results

Table 1.3 displays the OLS results. Each coefficient is the result of a separate regression of the outcome listed on the left on the level of CS137 in May 1986 and the controls listed at the bottom. The outcomes are standardized to mean zero and standard deviation one. A coefficient of $\widehat{\beta} = -0.01$ means that an increase in CS137 by 1 kBq/m^2 is associated with a decrease in the respective test score by 1 percent of a standard deviation. To facilitate the interpretation, we discuss the effect sizes relative to an increase in CS137 by one standard deviation ($sd(CS137) = 5.87$).

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

The movement of the coefficients in Columns (1)-(3) of Table 1.3 may at first appear surprising. The balancing tests in Table 1.2 show that people with lower education as well as people with less-educated parents lived in areas that received more fallout. If education had a positive effect on cognitive test scores, one would expect a negative correlation between CS137 and test scores. Once own and parental education is controlled for, one would expect the coefficient to become larger, i.e. less negative. However, in the German context it is plausible that we do not see the expected negative raw correlation. An important omitted variable here is the federal German education system, which varies in quality and institutions between states. The southern states of Bavaria and Baden-Württemberg traditionally had lower numbers of people with upper secondary or tertiary education, while at the same time their 9-th graders consistently achieved the highest standardized PISA test scores within Germany. Because both states have the highest average altitude and rainfall within Germany, they also received the largest amount of fallout. The weak correlation in Column (1) can be the consequence of those two counteracting forces. The heterogeneity in education systems can also explain the coefficient movement between Column (2) and (3). If states with the highest altitude and rainfall levels offer the best education and if altitude and rainfall are positively related with the level of fallout, then the coefficient becomes more negative when we control for both variables.

Table 1.3: OLS results: the effect of radiation on cognitive skills

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

Notes: This table displays the main estimation results. Each coefficient is the result of a separate OLS regression of the outcomes listed on the left on the ground deposition of CS137 in kBq/m^2*, controlling for the variables indicated below. In Columns (1)-(4), the test scores have been standardized. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels:* $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Non-linear Effects. In Appendix 1.A.4, we test for a non-linear dose-response relationship. Based on a polynomial regression, a spline regression and a level-log regression, we find little evidence of non-linear effects.

The Effect of Average Exposure. The estimates in Table 1.3 measure the effect of a higher initial potential exposure to radiation on cognitive skills. They are to be interpreted as reduced-form effects that summarize many channels through which cognitive skills are affected. However, the actual exposure over 25 years may be affected by at least two factors. First, due to the decay, the variation in radiation decreases over time. In addition, people who moved to an area with different radiation levels are exposed differently compared to people who stayed in their place of residence in 1986. In Appendix 1.A.4 we re-estimate the model in Column (4) of Table 1.3 but use as main regressors two measures of average exposure. The results are similar to the effects of the initial exposure. An increase in average exposure reduces cognitive test scores by between 3.2 and 9.6 percent of a standard deviation.

Robustness Checks. In Appendices 1.A.2 and 1.A.5, we assess the robustness of our estimation and inference. We show that the results are robust to different linking procedures of the radiation and survey data (1.A.2). We also rule out that the results are driven by selective mortality (1.A.2), design-based attrition or missing survey information (1.A.2). To account for multiple hypothesis testing, we perform a step-down correction of the p-values as well as a summary index test. Furthermore, we relax the parametric assumption of the normality of the error term and calculate the standard errors based on randomization inference. Finally, to allow for cross-sectional dependence within states, we perform a cluster bootstrap-t procedure (Cameron et al., 2008). Our inference is robust to these refinements of the calculation of the standard errors.

1.6.2 IV estimates

In Table 1.4, we present the instrumental variable estimates from regressions with all controls and state fixed effects. In Appendix 1.A.3, we report the first-stage, reduced-form and second-stage results with varying sets of controls. Column (1) reproduces the OLS results from Column (4) in Table 1.3, whereas Columns (2) and (3) report the first-stage coefficients and F-statistics of the IV estimation.

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

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

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

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

Finally, the difference can be explained by heterogeneous treatment effects. If treatment effects are not constant across the population, the IV estimator identifies a local average treatment effect (LATE). With a continuous instrument, this means that estimator places a larger weight on municipalities with a higher level of compliance. In Appendix 1.A.3, we explore potential sources of heterogeneous treatment effects. We find a stronger first-stage relationship in municipalities in southern states and in municipalities at above-median altitude. We also test for differences between urban and rural areas but find no significant difference. This suggests that the IV estimator places a larger weight on areas in the south and at higher altitude, which may deliver larger estimates than the OLS estimator, whose weights are determined by the state fixed effects.

Table 1.4: IV results: the effect of radiation on cognitive skills

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

Notes: This table displays the IV results. Column (1) reproduces the OLS results from Column (4) in Table 1.3. Columns (2) and reports the first-stage coefficients of CS137 regressed on the instrument and all controls mentioned in section 1.5.1 and the corresponding F-statistics, respectively. Column (4) reports the reduced-form coefficients of separate regressions of the outcomes on the instrument and controls. The main IV results are displayed in Column (5). Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

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

The estimates presented in Tables 1.3 and 1.4 show that the radiation induced by Chernobyl had significant negative effects on cognitive performance. A one-standard-deviation increase in ground deposition reduces cognitive test scores between 4 percent and 8 percent of a standard deviation. With one percent of a standard deviation being roughly equivalent to the cognitive skills acquired in one school year, this means that receiving this additional radiation dose reduces a person's human capital by the equivalent of 4-8 percent of a school year.[19]

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

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

The closest study for comparison is Almond et al. (2009). While their main specification is semi-parametric and, thus, difficult to compare, they also use the log level of fallout in some regressions. In Table 1.A.4.1, we estimate a similar specification and find that an increase in radiation by 100 log points reduced test scores by 7.6 percent of a standard deviation. The results in Almond et al. (2009) are significantly larger. They report an increase in math

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

scores by almost 100 percent and an increase in overall GPA by 67.5 percent of a standard deviation. The effects found by Black et al. (2013) for Norway in the 1950s are considerably smaller. They report an effect between 2 percent and around 25 percent of a standard deviation.[20] One reason why these effects may differ is the fact that the contamination in Norway in the 1950s was only half of that in Sweden in the 1980s.

These comparisons suggest that the effects of post-natal exposure are an order of magnitude smaller than the effects of exposure during pregnancy. If cells are damaged while crucial body functions develop, it is unsurprising that they have larger effects than after birth, when this process has been finished. Nonetheless, our effects are economically significant, not least due to the relative number of people exposed after birth. In West Germany in the 1980s, the number of people between weeks 8 and 25 of gestation at any point in time was around 200,000. On the contrary, the size of the birth cohorts 1956-1985 was 24 million.[21] Therefore, the in-utero studies document a very large effect of an environmental shock on a small number of people, whereas our paper documents a smaller effect but for a population that is over 100 times larger.

1.7 Additional Results

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

1.7.1 Heterogeneous Effects

In Table 1.5, we explore whether the impact of radiation exposure on cognitive skills differs between demographic groups. For each set of groups, we estimate full interaction models that interact the ground deposition of CS137 with mutually exclusive dummies for each group. For example, in Column (1), we interact the ground deposition with a dummy for male and a dummy for female, which provides with separate estimates for both groups.[22]

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

[21] Source: vital statistics provided by Destatis.

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

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, 2013), 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 *post-natal* exposure to radiation, those on brain cells and vital organs, are more likely to manifest themselves at older ages. However, because the youngest cohort was only 25 years old when they took the cognitive skills tests, we cannot observe what their test scores will be at age 50. Another potential explanation is that parents with young kids in 1986 particularly tried to shield their children away, thereby reducing the impact on later-life outcomes.

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

Finally, in Column (4), we assess if the effects differ between people who, in 1986, lived in the GDR versus West Germany. Unlike in West Germany, the population in the GDR received little information about the disaster and its likely consequences, and was even encouraged to consume foods that were potentially contaminated. Given these differences,

indicators in the regression. With only one indicator included, in this case a female dummy, the parameters are identified.

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

Table 1.5: Heterogeneous effects

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

Notes: Each column reports the result from a regression of the standardized cognitive skills index on a full interaction between the ground deposition of CS137 and mutually exclusive group indicators. In all regressions, we control for individual and municipality characteristics, as well as state fixed effects. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

1.7.2 Evidence on Cognitive Decline

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

Effects across age groups. First, we provide more evidence on heterogeneous effects across age groups. For this purpose, we estimate a difference-in-difference specification whereby we interact seven mutually exclusive dummies for people's age in May 1986 with the level of CS137,

$$y_{itms} = \sum_{t=1}^{7} \gamma_t \times CS137_{ms} + \delta_t + \boldsymbol{X'_{im}}\gamma + \delta_s + \varepsilon_{itms}. \tag{1.5}$$

The coefficients γ_t measure the effect of an increase in CS137 by 1 kBq/m^2 on the cognitive skill index for the seven age groups $t = 1, \ldots, 7$, controlling for average differences across age groups (δ_t), state fixed effects (δ_t) and the same individual, municipality and county characteristics as before.

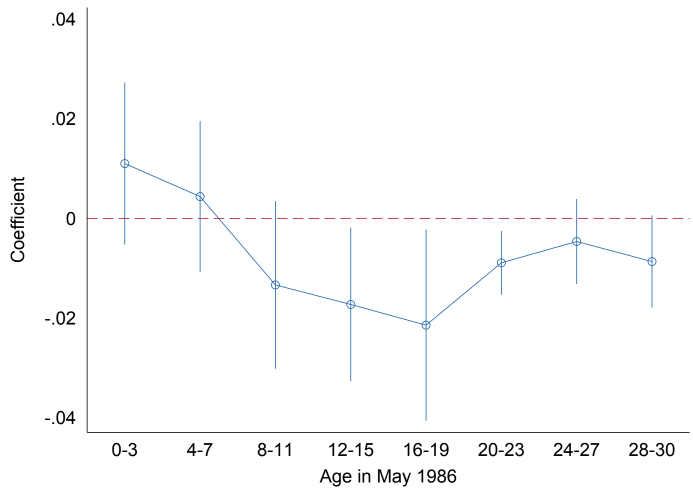


Figure 1.3: Differential effects by age group

Notes: This graph plots the estimated coefficients $\widehat{\gamma_t}$ and 95 percent confidence intervals for the age-specific effect of CS137 on the cognitive skills index. The estimates are based on Equation (1.5). Standard errors are clustered at the county level.

Figure 1.3 displays the estimated coefficients $\widehat{\gamma_t}$. While we find no significant effects for people aged 0-7 in May 1986, we find strong negative and in most cases statistically significant effects for ages 8 and older. The effects are strongest for people first exposed at ages 16-19. This can be seen as evidence that exposure matters most for cognitive skills during puberty, a time of significant re-modeling of the brain. However, we lack the statistical power to identify significant differences between age groups, such that the evidence is to be seen as suggestive.

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

Table 1.6 displays the estimated effects of an increase in radiation by 1 kBq/m^2 on the log number of diagnoses. The effect on dementia is positive once we control for geographic characteristics. In Column (3), when we condition on state fixed effects, the estimates become significant at the 5 percent-level. The point estimate of -0.077 means that an increase in radiation by one standard deviation increases the incidence of dementia by 45 percent. While this number sounds drastic, it has to be seen in relation to the mean of 148 diagnoses over an eleven-year period. An increase by 45 percent would mean an increase by 0.6 diagnoses per hospital per year. In Column (4) we use the same instrument as in the previous analysis. The first-stage relationship is different than in Table 1.4 because not all municipalities have a hospital. The IV estimate is larger than the OLS estimate. The point estimate of -0.12 is equivalent to a 70 percent-increase in the number of dementia cases for a one-standard-deviation increase in radiation. The results on asthma and injuries suggest that our estimates are unlikely to reflect a general health effect or some spurious relationship.

1.7.3 Evidence on behavioral responses

In Appendix 1.A.4, we test for behavioral responses along three margins, namely internal migration, employment and education. We run the same regressions as in (1.3), using as outcomes a binary indicator whether a person has moved until a given year, an indicator whether a person was employed in a given year and information on the number of years of completed education. We find evidence on neither of them. The effects on migration, employment and years of education are close to zero and precisely estimated. The only significant effect we find is on hours in continuing education, education people pursue while being employed. An increase in radiation by one standard deviation decreases the time spent in continuing education in 2010 by 9 hours (6.7 percent of the mean). Taken together, these results show little evidence that Chernobyl led to fundamental changes along these three margins.

Table 1.6: Hospital diagnoses, 2006-2016

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

Notes: This table displays the estimation results for the effect of average ground deposition of CS137 on the log cumulative number of hospital diagnoses between 2006-2016 at the municipality level. Each coefficient is the result of a separate regression of the variables on the left on the level of CS137 in May 1986 and the controls listed at the bottom. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

1.8 Conclusion

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

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

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

This chapter 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.

APPENDIX

1.A.1 Data Description

Sampling in the ALWA subsample.

As described in section 1.4, 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 percent out of all sampled addresses, and 45.9 percent 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.

Competence tests

Further details on test scores

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

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

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

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

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

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

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

Scientific literacy. The concept of *scientific literacy* follows the concept of the American Association of Advancement of Science (AAAS) and PISA.

Based on 22 items, this tests describes individual knowledge of basic scientific concepts and facts (KOS), divided into the content-related components *matter*, *systems*, *development* and *interactions*, and the understanding of scientific processes (KAS), divided into the process-related components *scientific enquiry* and *scientific explanations*, which are required for personal decision-making. The maximum attainable test score is 28 points. The maximum processing duration is 25 minutes by paper-pencil questionnaire.

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

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

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

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

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

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

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

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.

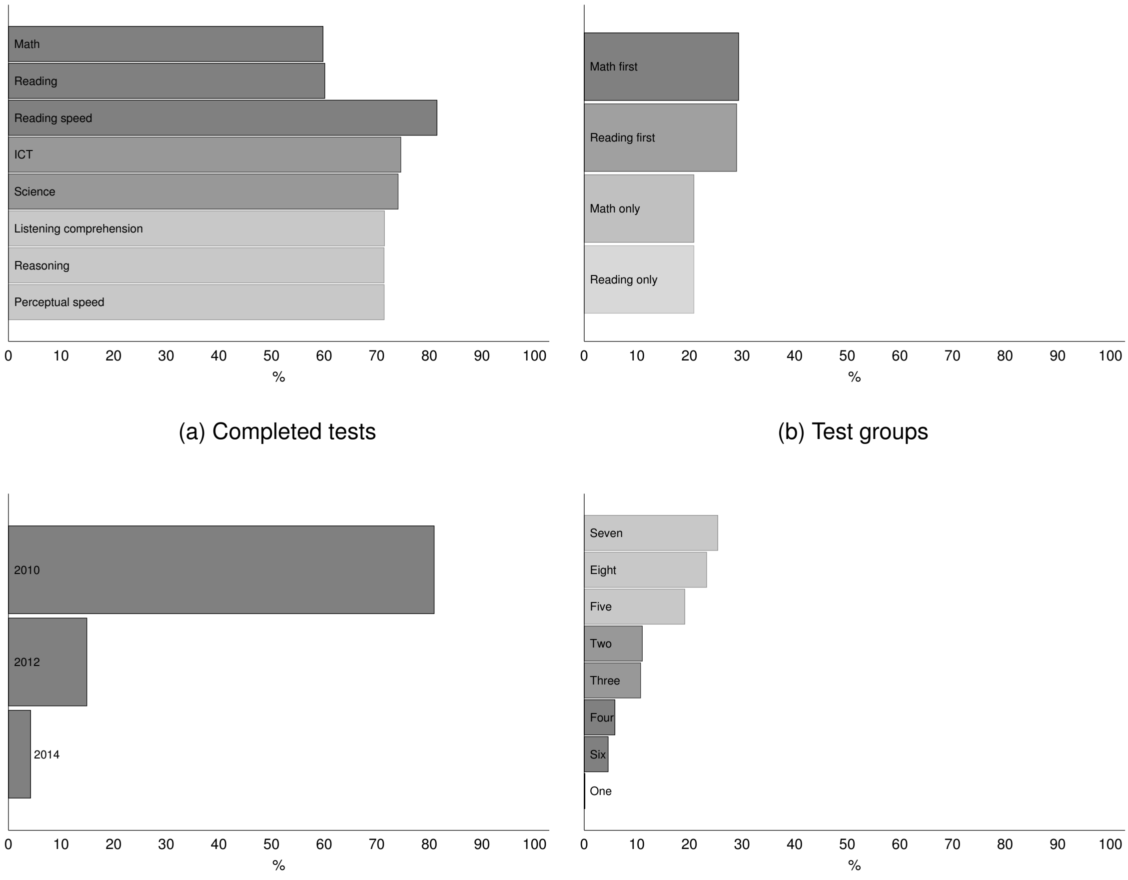
Besides an overall cognitive skill index that contains all eight test scores, we construct sub-indices for skills based on crystallized and fluid intelligence (Cattell, 1987). Research in radiobiology shows that radiation exposure has a larger effect on crystallized rather than fluid intelligence (Squire, 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.

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 1.A.1.1 c, some only started in the second and some few only in the third period.

As shown in figure 1.A.1.1 b 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 1.A.1.1 a 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



(a) Completed tests

(b) Test groups

(c) Year of first test

(d) Number of tests completed

Figure 1.A.1.1 : Participation in cognitive tests

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

test). Overall, 4,423 participants performed at least one test. Figure 1.A.1.1 d 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 1.A.2, we test whether the non-participation in the competence tests is systematically linked to the level of radiation, which is not the case.

Regressor of interest and control variables

Table 1.A.1.1: Variables and Data Sources

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

continued

Table 1.A.1.1 continued

Variable	Source
Upper secondary	Dummy variable of participants' educational achievements in April 1986 who are not enrolled but older than six years: 1) Upper secondary 0) Others
Tertiary	Dummy variable of participants' educational achievements in April 1986 who are not enrolled but older than six years: 1) Tertiary 0) Others
In school or college education	Dummy variable of participants' educational activity in April 1986 who are older than six years: 1) Enrolled 0) Not enrolled
No degree	Dummy variable of participants' educational achievement in April 1986 who are older than six years and enrolled: 1) No degree 0) Others
Already attained lower secondary, secondary	Dummy variable of participants' educational achievement in April 1986 who are older than six years and enrolled: 1) Lower secondary, secondary degree 0) Others
Upper secondary	Dummy variable of participants' educational achievement in April 1986 who are older than six years and enrolled: 1) Upper secondary 0) Others

continued

Table 1.A.1.1 continued

Variable	Source
B – Municipality-level Variables	
Caesium137 kBq/m^2 (01. May 1986)	Continuous variable of the ground radiation of Caesium137 kBq/m^2 at the municipality of residence in 01. May 1986. We computed this variable for the municipality centroid based on the inverse-distance weighted average of the four closest measuring points. Source: Federal Office for Radiation Protection
Average Caesium137 kBq/m^2 (until 2010, decay corrected)	Continuous variable of decay corrected Caesium137 kBq/m^2 levels at the municipality of residence between 1986 and 2010. Decay formula: $CS137_t = CS137_0 \times e^{-0.024t}$, Source: Federal Office for Radiation Protection
Precipitation mm/m^2 (yearly average, 1981-1985)	Continuous variable of precipitation in mm/m^2, computed for the centroid of a municipality based on the inverse-distance weighted average of the four closest measuring points. Source: German Meteorological Service
Altitude in meter	Continuous variable of the municipality center's altitude. Source: Federal Agency for Cartography and Geodesy
Population/1000	Continuous variable of the municipalities' population in 1997 (in 1000). Source: Federal Agency for Cartography and Geodesy

continued

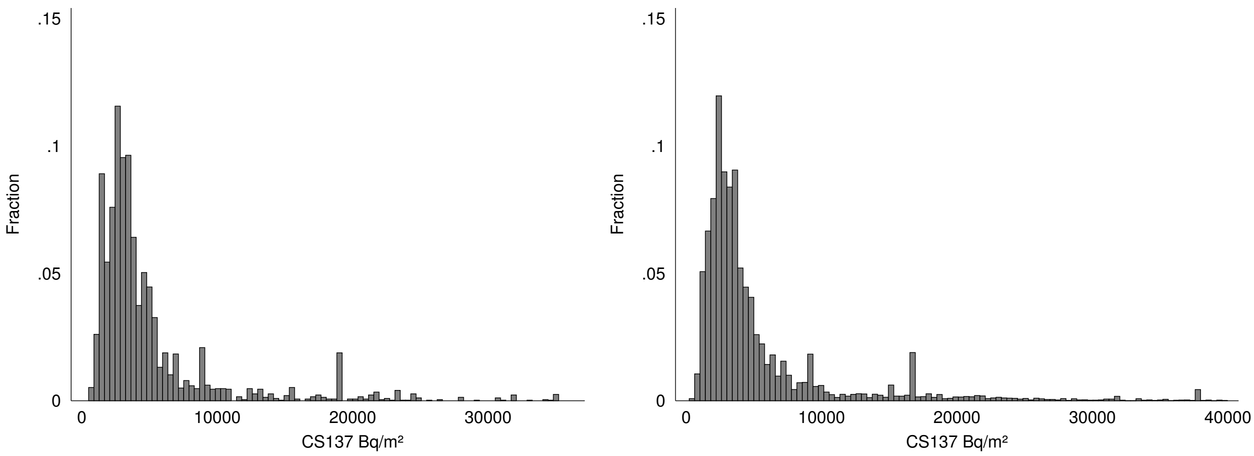
Table 1.A.1.1 continued

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

Notes: This table describes all control variables in our main specification in Column (4) of Table 1.3.

Distribution of fallout

Figure 1.A.1.2 displays the distribution of the fallout in our sample as well as the German population. Based on participants' municipality of residence in May 1986, Panel (a) displays the the ground deposition of CS137 in Bq/m^2 for the sample. Panel (b) shows the corresponding distribution for the entire German population, which we obtain by weighting the ground deposition in each municipality with the population in 1997. This was the first year for which consistent population data are available for the municipalities based on the same definition as the one used by the NEPS.



(a) Sample

(b) Population

Figure 1.A.1.2 : 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.

1.A.2 Robustness Checks

Goodness of fit for OLS regressions

Table 1.A.2.1 displays the adjusted R^2 for the OLS regressions in Table 1.3.

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

Table 1.A.2.1: Adjusted R^2 for main results.

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

Notes: This table displays the adjusted R^2 for the baseline results presented in Columns (1)-(4) in Table 1.3.

radiation data with the survey data based on assumptions. While we have fine-grained data on CS137 at a 3x3km grid-cell level, we only know a person's municipality of residence rather than the precise coordinates of their place of residence. In addition, the cell-level data have been generated by 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 1.A.2.2, we re-estimate the baseline model from Table 1.3, Column (8), with regressors based on different data linkages. The results in Table 1.A.2.2 strongly reject the notion that the results are driven our choice of linkage procedure.

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

closest measuring points, linked via the population center of a municipality[23]

- 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 municipality[24]
- Column (6): based on the radiation at the closest measuring point, linked via the population mode of a municipality
- Column (7): based on the unweighted average radiation in the entire municipality
- Column (8): based on the population-weighted average radiation in the entire municipality[25]

Testing for selective mortality

One important potential source of sample selection is selective mortality. Simply put, if radiation led to higher mortality among certain parts of the population, this population would be under-represented in our sample. To assess the importance of selective mortality, we obtained data on annual cohort-specific mortality data at the county level from the life tables of the German Statistics Office (Destatis).[26]. We run the following regression:

$$m_{crst} = \alpha + \rho_{rt}\ CS137_{crs} + \boldsymbol{X'_{cs}}\boldsymbol{\kappa} + \delta_s + \varepsilon_{crst}. \tag{B.1}$$

The number of deaths m_{cst} of age cohort r in county c state s in year t is regressed on the level of ground deposition of CS137 in May 1986 in the same county. To obtain the level of ground deposition for each county, we match the radiation data based on the county centroid. The vector of controls, $\boldsymbol{X_{cs}}$, includes county characteristics, namely the level of rainfall altitude at the centroid and the total population in the country. In addition, we control for state fixed effects δ_s. The error term ε_{cst} summarizes all determinants of mortality not captured by the regressors. The coefficient ρ_r measures the reduced-form effect of exposure to radiation in April 1986 on mortality between 1995 and 2010.

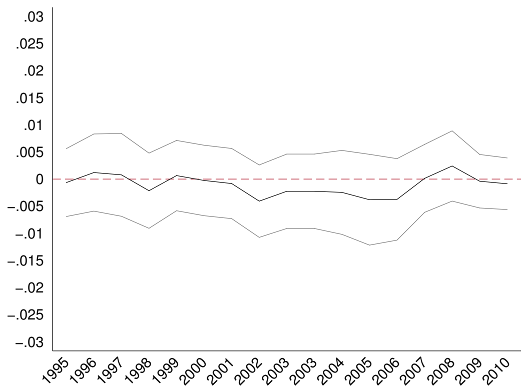
Figure 1.A.2.1 a displays the estimates ρ_{rt} for all cohorts, while the remaining Figures present cohort-specific estimates. We find no evidence that exposure to the Chernobyl fallout led to higher mortality until 2010.

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

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

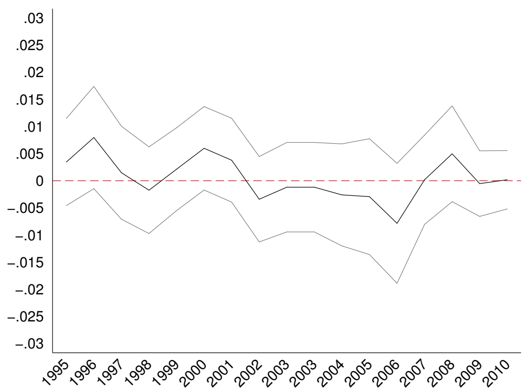
[25] The averages in Columns (7) and (8) were computed based on the 3x3km grid-level data. construct the population weights, we used night light data from 1996.

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

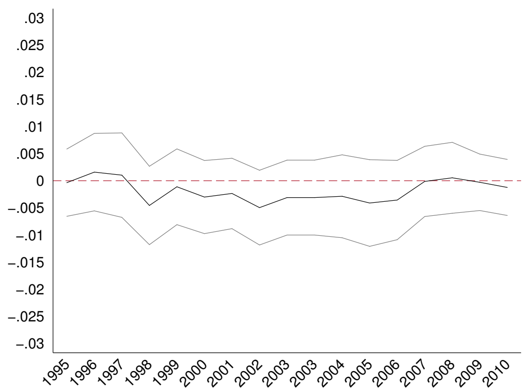


(a) All cohorts (1955-1985)

(b) Cohort 1975-1985



(c) Cohort 1965-1975



(d) Cohorts born before 1965

Figure 1.A.2.1 : Radiation exposure and mortality.

Notes: This graph displays the estimated effect of radiation exposure on standardized mortality in a given year. Both radiation and mortality vary at the county level. In all regressions, we control for county-level characteristics as well as state fixed effects. The lines in each panel represent the point estimates and 95 percent-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.*

Testing for design-based attrition

As shown in the descriptive statistics in Table 1.1, 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 1.A.2.3 shows, there is no evidence of systematic attrition or non-response once we add appropriate controls.

In Table 1.A.2.4, 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 1.A.1 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.

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 1.A.2.5 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 1.3, Column (8). The results show that, if anything, calculating the index based on all respondents leads to smaller estimates than calculating the index based on those who completed seven or eight tests.

Table 1.A.2.2: Robustness to the data linkage procedure

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

Notes: This table displays the estimation results whereby the regressor has been constructed with different data linkage procedures. See text in section 1.A.2 for a description of the linkage procedures. The controls are the same as in Table 1.3, Column (4). Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 1.A.2.3: Selection into competence tests

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

Notes: This table displays the results of separate regressions of dummy variables, indicating if an individual participated in the test or not, listed on the left on the ground deposition of CS137, controlling for the variables indicated at the bottom. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 1.A.2.4: Attrition

	(1)	(2)	(3)
A. Participation in competence test			
Cs137 kBq/m^2	-0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
(N)	5844	5844	5844
B. Missing personal information			
Cs137 kBq/m^2	0.001	0.000	-0.000
	(0.001)	(0.001)	(0.001)
(N)	4545	4545	4545
C. Municipality included in sample			
Cs137 kBq/m^2	0.0000	0.0004	-0.000
	(0.0002)	(0.0002)	(0.0002)
(N)	11197	11197	11197
Controls:			
County characteristics	No	Yes	Yes
Municipality characteristics	No	Yes	Yes
State FE	No	No	Yes

Notes: This table displays the results of regressions of indicators for participation or attrition on the level of fallout in 1986. In all regressions, we control for municipality characteristics and state fixed effects. In Panel A, the dependent variable is a binary indicator that equals unity if a person participated in the competence test. In Panel B, the dependent variable equals unity if the person is excluded from the estimation sample due to missing personal information. In Panel C, the dependent variable is an indicator that equals unity if a municipality was included in the NEPS SC6 sample and has at least one observation. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 1.A.2.5: The cognitive skills index with different definitions

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

Notes: This table displays the results of regressions of the standardized cognitive skills index on the level of ground deposition of CS137 and the controls listed at the bottom. In each row, we consider different sample definitions. In row one, the index is based on participants who completed all eight tests. In the second row, we consider all participants who completed at least seven tests, etc. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

1.A.3 Instrumental Variables: Details and Additional Results

First stages

Table 1.A.3.1 displays the first-stage coefficients for the regression Equation (1.4). Each coefficient is the result of a separate regression and is based on a different sample. The difference in samples is due to the fact that not every participant took every test, resulting in different first-stage coefficients and F-statistics. Column (1) displays the raw first-stage correlations without controls. In Column (2), we control for individual characteristics. In Columns (3) and (4), respectively, we add regional controls and state fixed effects. Introducing regional controls slightly reduces the size of the coefficients and the F-statistic decreases by about 50 percent. This indicates that part of the correlation of the instrument and the level of fallout is explained by regional factors such as population density, altitude and average rainfall. Nonetheless, even after introducing many controls, the F-statistic is above 30 in all regressions, which rules out a bias due to weak instruments.

Diagnostic tests in support of the exclusion restriction

In Table 1.A.3.2, we perform balancing tests by regressing individual pre-determined characteristics on the instrument and additional controls. Significant coefficients can be seen as evidence against the exclusion restriction as they suggest that the assignment of the instrument is not as good as random. In that case, it would be difficult to argue that the instrument is uncorrelated with the error term in Equation 1.3. The results in Table 1.A.3.2 suggest that the instrument passes this diagnostic test once we control for State fixed effects. The results support the identifying assumption that within states the assignment of the instrument is as good as random.

In Table 1.A.3.3, we perform an additional set of diagnostic tests by estimating the reduced form based on rainfall in different years. Ideally, we only want to find significant reduced-form effects based on rainfall in early May 1986 but not based on rainfall in early May in 1987 or 1988. Each coefficient is the result of a separate regression of the outcomes on the left on the instrument, individual-level controls, controls at the municipality- and county-level and state fixed effects. Column (1) displays the reduced form based on rainfall in 1986. All coefficients have the expected negative sign and 5 out of 11 coefficients are statistically significant at the 5 percent-level. In Columns (2) and (3) we estimate the same regressions but construct the instrument based on rainfall between May 1 and May 10, 1987 and 1988, respectively. Out of the 22 coefficients in both columns, one is significant at the 5 percent-level, which is consistent with random sampling variation. This indicates that the instrument works as it should. The assignment of the fallout is determined by rainfall while the plume was above Germany but not by rainfall on similar days in

Table 1.A.3.1: First-stage coefficients

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

Notes: This table displays the first stage estimation results for the effect of ln(Precipitation (mm/m^3) $\times$ Air contamination (mm/m^3)) on CS137. Each coefficient is the result of a separate regression of the variables on the left on a measure of ground deposition. Standard errors, clustered at the county level, are displayed in parentheses. The coefficients and F-Statistics differ because each estimation is based on a different sample owing to the fact that not every person took every test (see Appendix 1.A.1). Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 1.A.3.2: Balancing tests for IV

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

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

subsequent years.

Table 1.A.3.3: Diagnostic tests based on reduced form

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

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

Reduced-form and second-stage results

In Table 1.A.3.4, we report the reduced-form and second-stage results for different sets of controls. Each coefficient is the result of a separate regression of the outcomes on the left on the instrument (Columns (1)-(4)) or on the variation in CS137 that is predicted by the instrument (Columns (5)-(8)). Columns (1) and (5) display the results of univariate regressions. In Columns (2) and (6) we include individual-level characteristics. In Columns (3) and (7), we add municipality- and county-level controls, while in Columns (4) and (8) we add state fixed effects.

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

Further evidence of a LATE

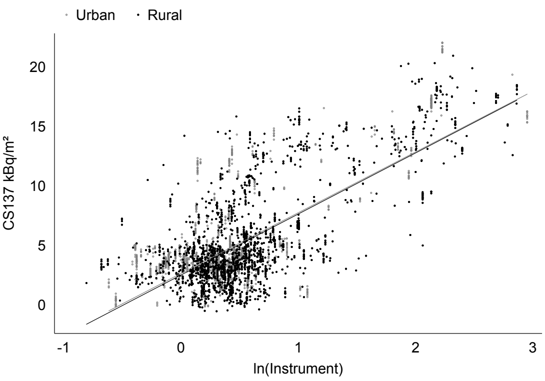
The analysis in section 2.3.3 reveals robust negative effects, although the IV estimates are significantly larger than the OLS estimates. One reason for this difference is that both estimators apply different weights to observations and, thus, identify different effects. Because our model includes state fixed effects, the weights of the OLS estimator are determined by the number of observations per state and the within-state variance in treatment (Gibbons et al., 2018). The IV estimator, in contrast, identifies the local average treatment effect. With a continuous treatment, the estimator places higher weight on municipalities in which the instrument has a stronger effect on radiation.

In Figure 1.A.3.1 , we provide evidence of differences in compliance between types of municipalities. Each panel plots the first-stage relationship with all controls and fixed effects for two distinct categories. In Panel a), we find similar first-stage relationships for areas with above- and below-median population density; both regression lines are virtually the same. In contrast, Panels b) and c) show significant differences. Panel b) plots the first stages for the southern states (Bavaria, Baden-Württemberg, Hesse and Rhineland-Palatinate) and the remaining states, labeled as 'north'. The first stage is strong in the south but not in the north. Similarly, Panel c) shows a stronger first-stage relationship for municipalities at above-median altitude. While not being an exhaustive list of categories, the three panels suggest it is plausible that the IV estimator identifies a LATE that gives larger weight to observations in southern states and those in municipalities at higher altitude.

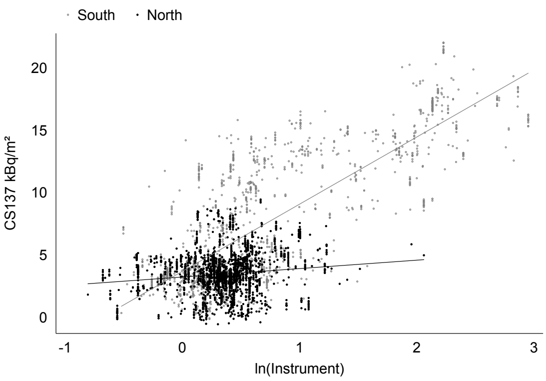
Table 1.A.3.4: IV second stage and reduced form

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

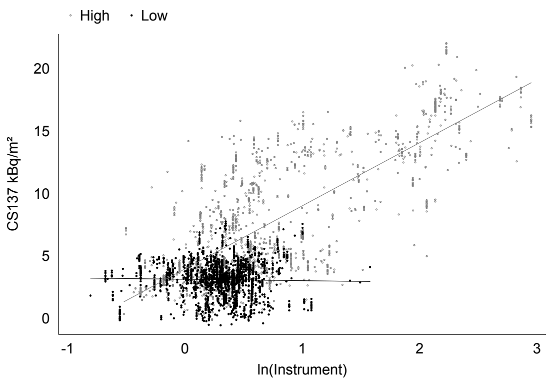
Notes: This table displays the reduced-form and second-stage estimations for the main specifications using ln(Precipitation (mm/m^3) $\times$ Air contamination (mm/m^3)) as instrument. In all regressions, we control for individual, municipality and county characteristics, as well as state fixed effects.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



(a) Urban vs. rural areas



(b) South vs. north



(c) High vs. low altitude

Figure 1.A.3.1 : First-stage correlations by category

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

1.A.4 Additional Results

Non-linear effects

In Table 1.A.4.1, we analyze if there is a non-linear dose-response relationship between radiation exposure and cognitive test scores. In each regression, the outcome is the cognitive skills index. For comparison, Column (1) reproduces the linear estimate reported in Column (4) of Table 1.3.

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

Table 1.A.4.1: Non-linear effects

	(1)	(2)	(3)	(4)
CS137 kBq/m^2	-0.008***	-0.015**		-0.021
	(0.003)	(0.006)		(0.034)
CS137 kBq/m^2 $\times$ CS137 kBq/m^2		0.000		
		(0.000)		
ln(CS137 Bq/m^2)			-0.076***	
			(0.029)	
CS137 kBq/m^2 $\times$ above median				0.014
				(0.034)
Controls:				
Individual characteristics	Yes	Yes	Yes	Yes
County characteristics	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Observations	4440	4440	4440	4440
Adj R^2	0.22	0.22	0.22	0.22

Notes: This table displays the estimates from OLS regressions of the standardized cognitive skill index on several functional forms of the ground deposition of CS137 as well as the control variables listed at the bottom. See section 1.5 for a detailed list of control variables. In Column (4), the ground deposition of CS137 is interacted with an indicator that equals unity if a person lived in May 1986 in an area with an above-median ground deposition. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

the ground deposition with a binary indicator that equals unity if a person lived in 1986 in an area with above-median ground deposition. While the point estimate is larger for people living in areas with above-median ground deposition, the coefficient is statistically insignificant, such that a linear relationship cannot be rejected. In Column (3), we impose a log-linear relationship, for which we find a large and statistically significant coefficient. For a one-standard-deviation increase in the log ground deposition (sd=0.72), we find a decrease in cognitive test scores by 5.6 percent of a standard deviation, which is similar to the estimate from the linear level-level model in Column (1).

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

The effect of average exposure

The coefficients in Table 1.3 measure the total effect of a higher ground deposition in 1986 on test scores 25 years later. The advantage of these estimates is that the initial level of ground deposition is arguably exogenous, such that these effects provide strong evidence of a negative causal effect of radiation on test scores. However, given that people were constantly exposed to the Chernobyl-induced radiation between 1986 and 2010, the interpretation of the magnitude is not straightforward. The most relevant measure for the regressor would be the cumulative effective radiation dose a person received during this period, namely the dose of radiation absorbed by all organs and tissues in a respondent's body. Obviously, such data would be very difficult to obtain as it would require measuring the energy absorbed by a person's tissue for every person in the sample.[27]

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 1.A.4.2, we report the regression results for different measures of average exposure. All outcomes are standardized to mean zero and standard deviation one. As a benchmark, Column (1) reproduces the main regression results with the ground deposition

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

in 1986 as the regressor. For years after 1986, we calculate the ground radiation based on the decay of CS137, $CS137_{mt} = CS137_{m0} \times e^{-0.024t}$ In Column (2), the regressor of interest is the average ground deposition of CS137 in a respondent's municipality of residence in 1986. In Column (3), we take into account internal migration and use as regressor the average ground deposition in a respondent's municipalities of residence between 1986 and 2010. Due to the constant decay of CS137, the variation in ground deposition across municipalities becomes smaller over time. As a result, the standard deviation of the average exposure is smaller than that of the initial exposure.

In Column (2), an increase in average ground deposition by 1kBq reduces the test scores by between 0.1 percent and 1.8 percent of a standard deviation. For the overall cognitive skill index, the effect is -1.1 percent 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 percent 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 percent 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 percent of a standard deviation, while the effect on the cognitive skills index is -3.2 percent of a standard deviation. In sum, these estimates have a similar magnitude to the reduced-form estimates in Column (1).

In Columns (4) and (5), we address three potential problems with our regressor in Column (3), the average ground deposition experienced by each person. The first problem is that people may move endogenously to avoid a higher radiation, which would bias the estimates. The second problem is measurement error in the regressor. We can only compute the amount of CS137 in a given year based on its decay, but we do not observe the extent to which the radioactive matter is washed into deeper layers of soil. Because a person's exposure is higher the closer the matter is to the surface, our way of computing the average ground deposition inevitably introduces measurement error, which, if unsystematic, biases the results towards zero. A third problem is non-random sorting into areas, which may be correlated with the level of fallout. We address these problems by instrumenting the average ground deposition between 1986 and 2010 with the initial ground deposition in May 1986 (Column (4)) and with the instrument described in section 1.5 (Column (5)). In both cases, the first-stage coefficient has the expected positive sign, suggesting that, on average, people who lived in areas with a higher ground deposition in 1986 were exposed to a higher average ground deposition over the following 25 years. The correlation is not perfect because people moved between places with different ground deposition. The instrumental variable estimates in both columns are considerably larger than the OLS estimates in Column (3). An increase in the average ground deposition by one standard deviation reduces cognitive test scores by 9.6 percent of a standard deviation. As described in section 1.6.2, the IV estimates may be larger because of measurement error, unobserved heterogeneity and because the IV identifies a local average treatment effect. As shown in Appendix 1.A.3, the IV estimator gives a higher weight to municipalities in the south than in the north and in general at higher altitude.

Evidence on behavioral responses

In Table 1.A.4.3, we explore the importance of several behavioral responses, namely internal migration, labor supply and investment in education. However, in our analysis we are constrained by the information available in our dataset. While the NEPS SC6 has rich information on some channels, we are unable to study several other behavioral responses such as changes in health behaviors, diet or exercise habits.

In the first panel of Table 1.A.4.3, 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 percent of the mean. Besides that, we find little evidence of the behavioral responses that we are able to measure.

Table 1.A.4.2: The Effect of Average Exposure, 1986-2010

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)
Math	-0.011***	-0.015***	-0.014**	-0.026***	-0.049**
	(0.003)	(0.004)	(0.006)	(0.008)	(0.021)
Reading	-0.014**	-0.018**	-0.012*	-0.034***	-0.064***
	(0.005)	(0.007)	(0.007)	(0.012)	(0.016)
Listening comprehension	-0.009**	-0.012**	-0.012*	-0.023**	-0.026
	(0.004)	(0.005)	(0.006)	(0.009)	(0.018)
ICT	-0.005	-0.007	-0.003	-0.012	-0.011
	(0.004)	(0.005)	(0.005)	(0.008)	(0.014)
Scientific literacy	-0.003	-0.005	-0.001	-0.008	-0.014
	(0.003)	(0.004)	(0.005)	(0.008)	(0.014)
Reasoning	-0.001	-0.001	-0.000	-0.002	-0.016
	(0.004)	(0.006)	(0.007)	(0.011)	(0.017)
Reading speed	-0.008**	-0.010**	-0.019***	-0.018**	-0.037**
	(0.004)	(0.005)	(0.007)	(0.008)	(0.018)
Perceptual speed	-0.004	-0.005	-0.007	-0.010	-0.004
	(0.003)	(0.004)	(0.006)	(0.008)	(0.017)
B. Indices					
Cognitive skill index	-0.008***	-0.011***	-0.010*	-0.019***	-0.030**
	(0.003)	(0.004)	(0.005)	(0.007)	(0.013)
Crystalized intelligence index	-0.009**	-0.011**	-0.008*	-0.020**	-0.034**
	(0.003)	(0.005)	(0.005)	(0.008)	(0.014)
Fluid intelligence index	-0.006*	-0.008*	-0.011	-0.014*	-0.018
	(0.003)	(0.004)	(0.007)	(0.008)	(0.014)
First-stage: dep. var. Cs137 kBq/m^2					
Cs137 kBq/m^2				0.424***	
				(0.006)	
Precipitation (mm/m^3) $\times$ Air contamination (mm/m^3)					3.732***
					(0.304)
F statistic				912.733	53.902
Controls:					
Individual characteristics	Yes	Yes	Yes	Yes	Yes
County characteristics	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Mean (CS137)	5.18	3.89	2.95	2.95	2.95
SD (CS137)	5.87	4.41	3.23	3.23	3.23

Notes: This table displays the estimation results for the effect of average ground deposition of CS137 from 1986-2010 on test scores. Each coefficient is the result of a separate regression of the variables on the left on a measure of ground deposition. Column (1) reproduces Column (4) in Table 1.3. In Column (2), the regressor is the decay-corrected average ground deposition from 1986 to 2010 in a respondent's municipality of residence in May 1986. In Columns (3) and (4), the regressor is the decay-corrected average ground deposition from 1986 to 2010, taking into account internal migration after 1986. In Column (4), we use the initial ground deposition in 1986 as an instrument for the average ground deposition between 1986 and 2010. In Column (5), we use the same instrument as in section 1.6.2. The mean and standard deviation of CS137 refer to the regressor used in each column. In all regressions, we control for individual, county and municipality characteristics, as well as state fixed effects. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 1.A.4.3: Evidence on behavioral responses

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

Notes: This table displays the results of separate regressions of the indicator variables listed on the left on the ground deposition of CS137. In all regressions, we control for individual and municipality characteristics, as well as state fixed effects. For migration the outcome is an indicator that equals unity if, until a given year, a person moved away from his/her municipality of residence in 1986. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

1.A.5 Inference

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 1.3 Columns (3) and (4) with the standardized cognitive skills index as dependent the variable. We repeat this procedure 10,000 times.

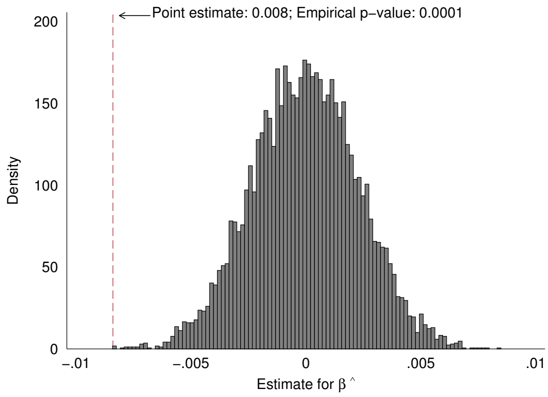
Figure 1.A.5.1 a displays the placebo distribution of 10,000 estimates with randomization across all observations, which allows us to assess the inference in a model without state fixed effects (Table 1.3 Column (3)). If the relationship was pure noise, a point estimate at least as extreme as -0.008 would be very unlikely to occur. In fact, in 10,000 replications, such a result only occurred once. The distribution in Figure 1.A.5.1 b corresponds to the estimations with state fixed effects presented in Table 1.3, 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 percent of all cases. This corresponds to an empirical p-value in a one-sided test of $p = 0.00026$.

Figure 1.A.5.1 c 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 1.3 in Column (3), only occurred 112 times, corresponding to an empirical p-value of $p = 0.0112$. The distribution in figure 1.A.5.1 d corresponds to the estimations with state fixed effects presented in Table 1.3, 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 $p = 0.0138$.

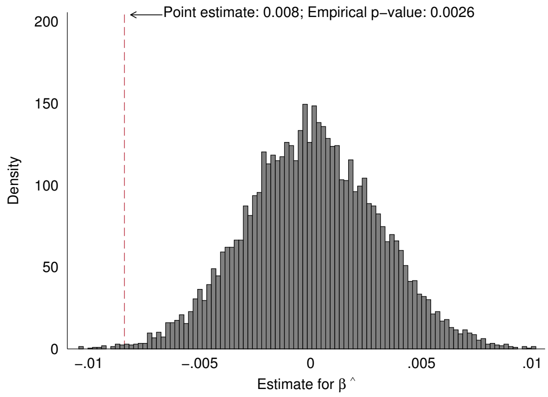
In sum, these results reinforce the conclusions drawn from our inference with clustered standard errors in section 2.3.3. If we consider the p-values of a two-sided hypothesis test, in which case the aforementioned p-values have to be multiplied by two, our estimates are statistically significant at the 5 percent-level.

Multiple hypothesis testing

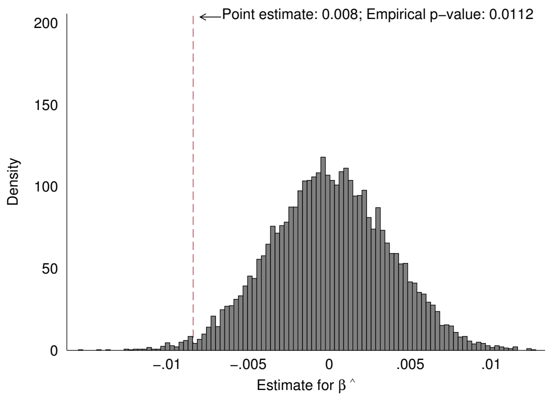
In our main analysis, we use eight cognitive test scores as outcome variables and estimate the impact of radiation on each outcome in separate regressions. However, because all of these outcomes represent different dimensions of the same latent factor cognitive skills,



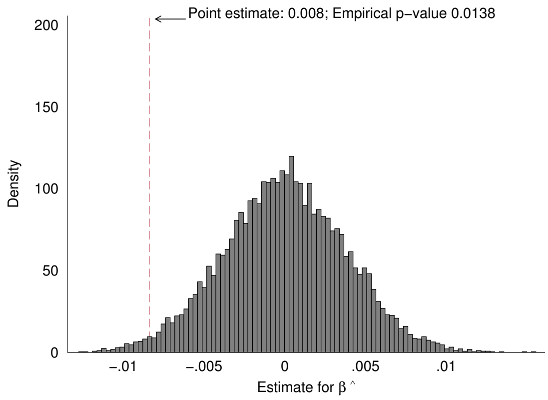
(a) Full randomization od CS137



(b) Randomization of CS137 within states



(c) Full rando. of cog. skill index



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

Figure 1.A.5.1 : 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 1.3 as well as the empirical p-values for one-sided tests.

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.

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

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

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

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 1.A.5.1 displays the p-values and q-values for all eight outcomes.[28] After the adjustment, three coefficients remain statistically significant at the 5 percent-level and one (reading speed) at the 10 percent-level.

Table 1.A.5.1: Q-values (p-values adjusted by step-down approach)

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

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

[28] Deviations of the q-values from Equation (E.1) are due to rounding.

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 1.A.5.2, the results only differ marginally between weighted and unweighted indices. Overall, these results, as well as those shown in Table 1.A.5.1, confirm the statistical significance of the negative effect of radiation exposure on cognitive skills.

Table 1.A.5.2: Summary index tests

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

Notes: This table displays the results of regressions of the indices listed on the left on the ground deposition of CS137 and the controls listed at the bottom. Column (1) reproduces the baseline results from Table 1.3 Column (4), whereby the standardized indices are unweighted.

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

Cameron et al. (2008) provide a bootstrap-based method that allows for the calculation of standard errors with few clusters. Rather than sampling single observations in a bootstrap sample, this procedure samples entire clusters. Table 1.A.5.3 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 1.3, the bootstrapped standard errors are larger, although most estimates remain statistically significant at the 5 percent- or 10 percent-level.

Table 1.A.5.3: Estimates with cluster-bootstrapped standard errors

	(1)	(2)	(3)	(4)
A. Individual test scores				
Math	0.003	0.002	-0.011***	-0.011***
	(0.004)	(0.005)	(0.004)	(0.004)
Reading	-0.001	0.001	-0.013***	-0.014***
	(0.008)	(0.015)	(0.003)	(0.003)
Listening comprehension	-0.003	-0.003	-0.008**	-0.009***
	(0.003)	(0.003)	(0.004)	(0.003)
ICT	0.000	0.001	-0.003	-0.005*
	(0.003)	(0.006)	(0.003)	(0.003)
Scientific literacy	0.001	0.002	-0.002	-0.003
	(0.001)	(0.004)	(0.003)	(0.003)
Reasoning	0.002	0.001	-0.001	-0.001
	(0.004)	(0.004)	(0.005)	(0.006)
Reading speed	-0.001	0.001	-0.010*	-0.008*
	(0.007)	(0.079)	(0.005)	(0.004)
Perceptual speed	0.003	0.003	-0.003	-0.004
	(0.004)	(0.004)	(0.003)	(0.003)
B. Indices				
Cognitive skill index	0.001	0.002	-0.007**	-0.008**
	(0.008)	(0.006)	(0.004)	(0.004)
Crystallized intelligence index	0.000	0.002	-0.007**	-0.009***
	(0.004)	(0.005)	(0.003)	(0.003)
Fluid intelligence index	0.002	0.003	-0.006	-0.006
	(0.012)	(0.009)	(0.005)	(0.004)
Controls:				
Individual characteristics	No	Yes	Yes	Yes
County characteristics	No	No	Yes	Yes
Municipality characteristics	No	No	Yes	Yes
State FE	No	No	No	Yes

Notes: This table corresponds to the main regression table 1.3. The standard errors in this table have been computed based on the wild cluster bootstrap-t procedure by Cameron et al. (2008). Significance levels: $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

1.A.6 Geographic information

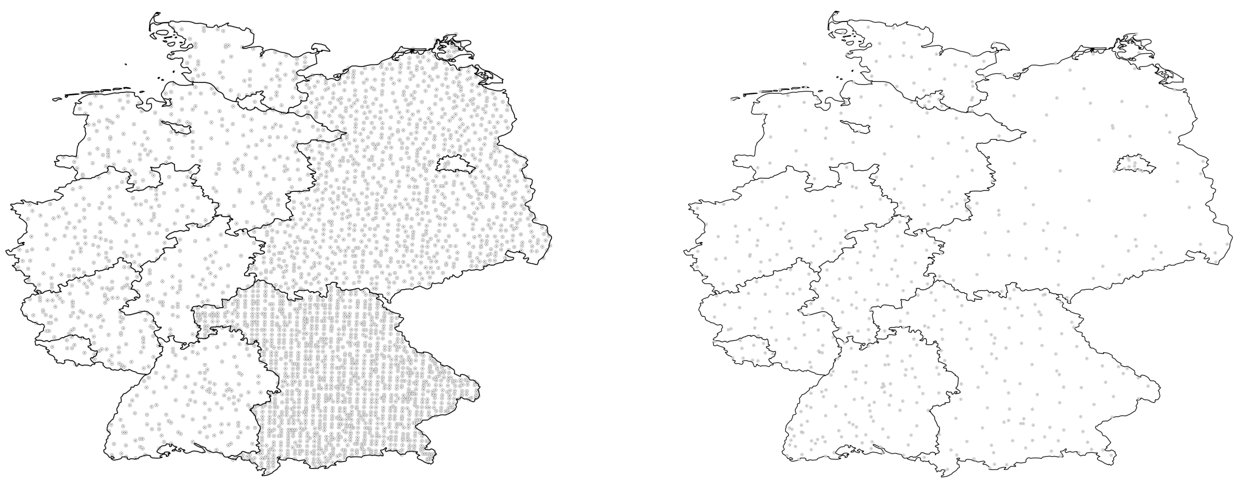
In this section, we provide further background information on the measurement of radiation and rainfall, as well as the climatic conditions in Germany around the Chernobyl disaster.

Measuring points for ground contamination. Figure 1.A.6.1 a shows the distribution of 3.448 measurement points for soil contamination, which is measured by an in-situ gamma ray spectrometer. Due the federal structure of Germany, several institutions were involved in the collection of measurements (Bavarian State Ministry for Regional Development and Environmental Issues; The Bavarian State Ministry for Food, Agriculture and Forestry; The Institute for Water, Soil and Air Hygiene of the Federal Health Office; State Office for Environmental Protection in Baden-Wuerttemberg; RWTH Aachen University). However, the leading institute was the Institute of Radiation Hygiene (ISH) of the former the German Federal Health Office (BGA) which coordinated, collected and evaluated measurements.

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

In the GDR the "Staatliche Amt fuer Atomsicherheit und Strahlenschutz" (SAAS) was the only institute responsible for the execution and evaluation of measurements. A country-wide measurement program was initiated with a 8x8 km grid (Bundesamt für Strahlenschutz, 2016). However, figure 1.A.6.1 a reveal that the measurement points in the GDR in our dataset is not as dense as in Bavaria. After the collapse of the GDR the Institute for Water, Soil and Air Hygiene of the Federal Health Office (WaBoLu) combined the data of in-situ gamma ray spectrometer collected by the GDR and the FRG to the dataset we are using. Only highly-reliable measurements were used by the WaBoLu, which explains missing measurements points in the GDR. In 1994 the WaBoLu was integrated in the Federal Environment Agency. The Federal Office for Radiation Protection provided us the radiation data which is the successor organization of the (ISH).

Measuring points for rainfall. Figure 1.A.6.1 b 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 1.A.6.1 a, a uniform distribution is evident across the county. The principal aim of this distribution is



(a) CS137 (b) Precipitation

Figure 1.A.6.1 : Measurement points

Notes: This figure displays the measurement points of CS137 and precipitation. Source: The German Meteorological Service, The Federal Office for Radiation Protection

the collection of weather data which is representative for the whole county. Furthermore, location requirements determine the exact distribution of weather stations. For example, the inclination of the surrounding terrain should not exceed a specific limit, operation near high buildings is not possible and measurement operation should be executable for at least ten years (Wetterdienst, 2017).

Trajectory of the radioactive plume. The radioactive plume reached Germany three days after the disaster, on April 29, 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 1.A.6.2 , which shows the air concentration of radioactive particles (radionuclides) in four measuring stations in different parts of Germany. The station Brotjacklriegl, a mountain in the south-east, close to the border with the Czech Republic and Austria, is located in the area 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 29 and May 8, 1986 is shown in Figure 1.A.6.3 a. Darker color represents higher precipitation. We determine this period as critical period based on our observations in Figure 1.A.6.2 .

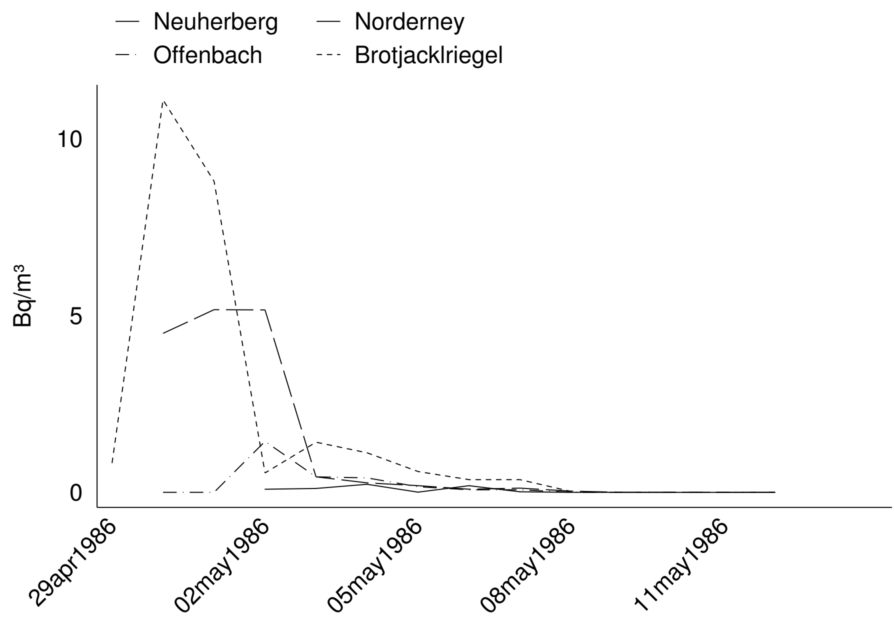
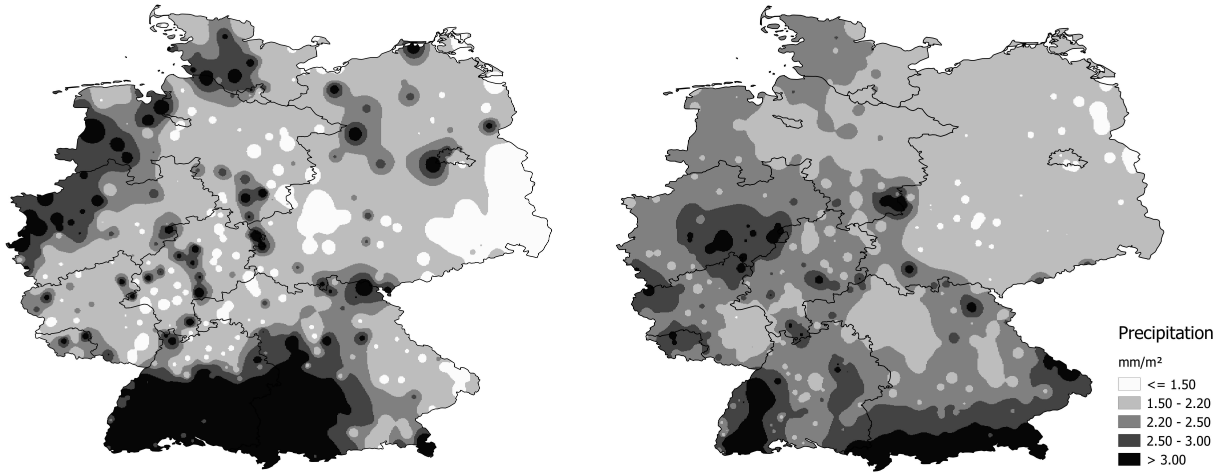


Figure 1.A.6.2 : Air concentration of radioactive particles in 1986

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).

Comparing the level of precipitation with the ground deposition of CS137 shown in Figure 1.1a, there appears to be a high correlation between the two. Figure 1.A.6.3 b, in contrast, shows the average precipitation between 1981 and 1985. A comparison of Figures 1.A.6.3 a and 1.A.6.3 b, 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.

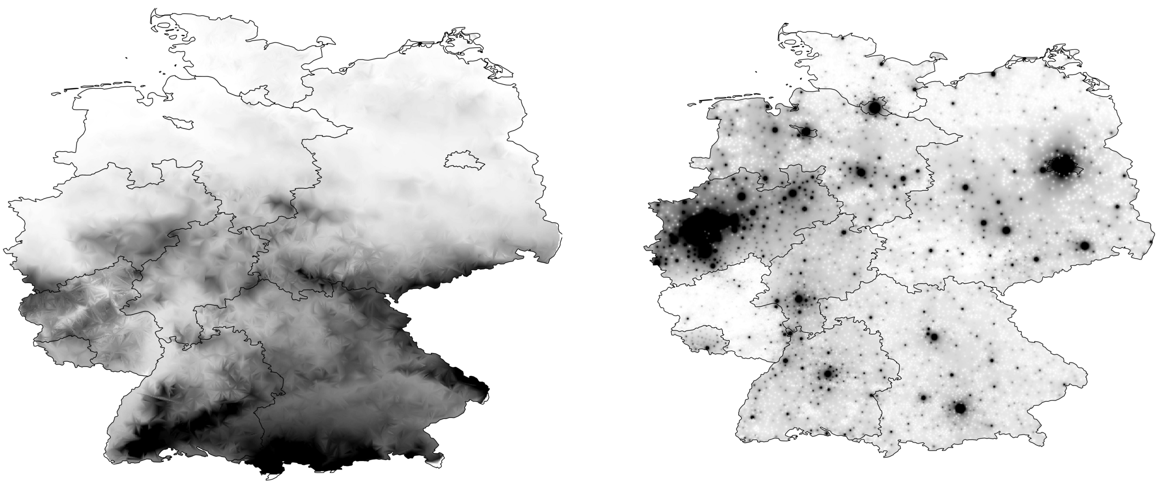
Altitude and population density. In the regressions, we control for altitude and population density, two potential determinants of both ground contamination and test scores. Figures 1.A.6.4 a and 1.A.6.4 b display the distribution of both variables across space.



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

Figure 1.A.6.3 : Average daily precipitation

Notes: This Figure displays average levels of precipitation. The dark regions correspond to higher levels. Source: The German Meteorological Service.



(a) Topography (b) Population density

Figure 1.A.6.4 : Altitutde and population density

Notes: This Figure displays altitude and population density. The dark regions correspond to higher levels. Source: Federal Agency for Cartography and Geodesy.

CHAPTER 2

Low Emission Zones for Better Health

This chapter studies health effects from restricting the access of high-emission vehicles to inner-cities by implementing Low Emission Zones. For identification, we exploit variation in the timing and the spatial distribution of the introduction of new Low Emission Zones across cities in Germany. We use detailed hospitalization data combined with geo-coded information on the coverage of Low Emission Zones. We find that Low Emission Zones significantly reduce levels of air pollution in urban areas and that these improvements in air quality translate into population health benefits. The number of diagnoses related to air pollution is significantly reduced for hospitals located within or in close proximity to a Low Emission Zone after it becomes effective. The results are mainly driven by reductions in chronic cardiovascular and respiratory diseases.[1]

[1]This chapter is based on joint work with Nico Pestel. A prior version was published as IZA Discussion Paper No. 12545

2.1 Introduction

Air pollution is a major concern for human health and well-being across the globe. According to the World Health Organization, about seven million premature deaths per year as well as a wide range of health hazards, in particular respiratory and cardiovascular diseases, can be attributed to poor air quality (WHO, 2018b).[2] While adverse health effects of air pollution may be more severe in the developing world, many places in high-income countries are also faced with serious violations of air quality standards. This creates large economic costs through hampered human capital formation (Graff Zivin and Neidell, 2013), increasing defensive medical spending (Deschênes et al., 2017) and reductions in workers' labor supply and productivity on the job (Graff Zivin and Neidell, 2018).

Emissions from traffic are a major source of ambient air pollution in densely populated urban areas (Karagulian et al., 2015). Automobile exhaust is particularly harmful to human health because it is mostly emitted close to the ground. Thus, reducing air pollution from traffic is of great importance for environmental policy-making. In the European Union, a key policy measure to reduce ambient air pollution in inner-cities is the implementation of Low Emission Zones, signposted areas where access of vehicles is regulated, typically banning high-emitting vehicles from entering the zone altogether. While access regulations impose costs on local residents and businesses, benefits may accrue in form of improved health, worker productivity and human capital. However, there is relatively little evidence about potential health benefits from policy interventions aiming at improving air quality in inner-cities. This is remarkable since policy measures, such as Low Emission Zones, are typically justified by improvements in population health.

In this chapter, we study whether the implementation of Low Emission Zones affect population health through improvements of air quality by evaluating the staggered introduction of this policy measure across German cities since 2007.[3] For causal identification of the health impact of Low Emission Zones, we exploit variation in the timing as well as the exact geographic coverage of Low Emission Zones across Germany in a difference-in-differences framework. The policy treatment of introducing a Low Emission Zone is triggered by local violations of European Union air quality standards. The decision to implement a Low Emission Zone is then forced upon cities by state governments who are responsible for compliance with air quality legislation. We exploit policy variation in the extent to which inner-city areas, usually the city center, are covered by Low Emission Zones across time as well as

[2]Air pollution is also the main cause of more than 440,000 deaths per year in Europe and 62,000 deaths in Germany alone (European Environmental Agency, 2018; Landrigan et al., 2017).

[3]Germany is currently the country which has established most Low Emission Zones based on relatively strict European Union legislation requiring legal actions against air quality standard violations. Low Emission Zones have been implemented in other European countries and will become more frequent in the near future. As of 2018, more than 200 Low Emission Zones have been established in European cities and this number will increase to more than 300 until 2025 (see Figure 2.A.1.1).

between and within cities.

We combine information on the geographic coverage of Low Emission Zones with rich panel data on the universe of German hospitals over the period from 2006 to 2016 with precise information on hospital locations and the annual frequency of detailed diagnoses based on international standard classification (ICD-10). We mainly focus on cardiovascular and respiratory diseases, which have been shown to be affected by key target pollutants like particulate matter and nitrogen oxides (Graff Zivin and Neidell, 2013). Additionally, we complement the analysis by looking at further outcomes related to infant health (low birth weight) as well as to outcomes potentially affected by reduced traffic within Low Emission Zones (injuries and stress). While it is straightforward to determine the distance of hospital locations to Low Emission Zones, the hospital data do not contain information on patients' residential locations which would allow us to assign the treatment to a hospital's potential pool of patients.[4] That is why we employ several approaches to construct hospitals' catchment areas, i.e., geographic areas in the surrounding of hospital locations from where admissions are likely to be from. Overlaying hospitals' catchment areas with Low Emission Zone coverage allows us to compute the share of hospital catchment areas treated by the policy. This means that we estimate reduced-form and intention-to-treat effects of Low Emission Zone introductions on hospital admissions. In order to establish that our estimates of Low Emission Zones' health impacts can indeed be attributed to improvements in local air quality ("first stage"), we additionally use data from Germany's official air pollution monitoring system and assign monitor locations to Low Emission Zones and test whether air pollution is affected by the coverage of a Low Emission Zone.

Our results show that Low Emission Zone introductions benefit population health. In a first step, we confirm that Low Emission Zones improve air quality, mainly by decreasing the frequency of exceeding regulatory thresholds. We do not find effects on traffic volume in- or outside the Low Emission Zone. In a second step, we show that these improvements in air quality translate into lower prevalence of several air pollution-related diagnoses, especially diseases of the circulatory and the respiratory system, among hospitals whose catchment areas are covered more by a Low Emission Zone. These results appear to be mainly driven by reductions in diagnoses of non-emergency diagnoses of chronic diseases and not so much by emergency cases. Low Emission Zones do not reduce the incidence of low birth weight significantly. Furthermore, we do not find significant effects on injuries or diagnoses of stress potentially related to changes in traffic volume.

The analysis presented in this chapter contributes to the literature in several ways. First, we add to the large literature on the causal impacts of air pollution on human health in both epidemioloy (Pope III, 2000; Pope III and Dockery, 2006) as well as in economics

[4] In Germany, access to individual-level administrative data on hospitalization with precise residential information is unfortunately not available.

(Graff Zivin and Neidell, 2013). Second, we contribute to a smaller number of studies in the economics literature evaluating the direct impact of Low Emission Zones on local air pollution. The two papers that are closest to ours are Wolff (2014) and Gehrsitz (2017) who both document significant drops in ambient air pollution after Low Emission Zone introductions in treated cities in Germany. A related literature in transportation research documents similar findings (Morfeld et al., 2014; Malina and Scheffler, 2015; Jiang et al., 2017). Wolff (2014) further shows that reductions in air pollution are driven by an improvement of the vehicle fleet in terms of emission standards. Our contribution is an extension of the analysis including the most recent Low Emission Zone implementations in Germany at a higher spatial accuracy using within-city variation. Third, the findings of our paper contribute to our understanding of the health benefits associated with policy measures regulating traffic in urban areas. Gehrsitz (2017) evaluates the effects of Low Emission Zones on infant health outcomes in Germany. However, the results do not indicate substantial reductions in the prevalence of low birth weight or the number of stillbirths in Germany following a ban of high-emission vehicles.

Simeonova et al. (2018) study the health effect of another policy measure to improve inner-city air quality, showing that implementing a congestion tax in central Stockholm reduced ambient air pollution and significantly decreased the rate of acute asthma attacks among young children. While children, especially newborns, are particularly vulnerable to detrimental environmental conditions (Almond and Currie, 2013), the elderly as well as the working-age population are also negatively affected by air pollution (Schlenker and Walker, 2016; Deschênes et al., 2017; Karlsson and Ziebarth, 2018). In this chapter, we are able to study the full range of diseases potentially affected by ambient air pollution among all age groups. Salvo et al. (2018) show that removing Diesel trucks from passing through the inner-city of São Paulo by inaugurating a beltway had positive effects on congestion, pollution, health and mortality benefiting the megacity's population. The results of our paper indicate that potential improvements in population health from reductions in traffic emissions are not restricted to locations starting from extremely high levels of air pollution but that health improvements can be achieved also for medium-sized cities with ex ante moderate levels of air pollution.

The remainder of this chapter is structured as follows. In section 2.2, we provide background information about German Low Emission Zones, targeted pollutants and show the effect of Low Emission Zones on air pollution. Section 2.3 describes the empirical analysis. Section 2.4 concludes.

2.2 Institutional Background and Data

2.2.1 Low Emission Zones in Germany

Air quality standards in Germany are determined by European Union (EU) legislation. Since the mid-1990s, the EU has established a legal framework in order to aspire levels of air quality that do not give rise to significant negative impacts on and risks to human health and the environment. The EU Directives 2008/50/EC and 1999/30/EC regulate measures to improve ambient air quality in all EU member states. The EU's legal framework has to be adopted by national law. It defines measurement procedures, limit values and alert thresholds for various target air pollutants in ambient air, among others nitrogen dioxide and particulate matter (see Table 2.A.1.1 in Appendix 2.A.1 for an overview). Violations of air quality standards require member states to adopt action plans with appropriate measures to reduce air pollution. Ultimately, non-compliance may result in penalty charges.[5]

In Germany, the 16 federal states are responsible for compliance with the EU air quality standards. In case of violations, state governments are obliged to develop city-specific Clean Air Plans (*Luftreinhaltepläne*), defining a bundle of measures aiming at lasting improvements of air quality in compliance with the EU standards. Usually, the respective city administrations as well as other stakeholders (e.g., business or environmental protection associations) are involved in the decision-making process. However, state governments ultimately decide on the Clean Air Plans and may overrule the views of local decision-makers and enforce the implementation or strictness of certain measures to be defined if they are deemed to be necessary to achieve compliance with the air quality standards. The implementation of a Low Emission Zone is the most tangible measure from the Clean Air Plan tool box to reduce traffic emissions in urban areas.[6]

Low Emission Zone implementations are controversially debated on the local level when they are announced for a given city. On the one hand, Low Emission Zones are unpopular as they impose restrictions on car owners and may create costs for local businesses. On the other hand, national environmental protection associations have filed a number of lawsuits aiming at implementing stricter measures to enhance compliance with the EU air quality standards more quickly, usually speeding up the adoption of Low Emission Zones or enforcing stricter regulations.[7] This means that, after there have been violations of air

[5]If a member state fails to adopt measures that are sufficient to reach the limit values in reasonable time, the EU can start an infringement procedures. In May 2018, there were 16 infringement cases pending against member states (Belgium, Bulgaria, the Czech Republic, Germany, Greece, Spain, France, Hungary, Italy, Latvia, Portugal, Poland, Romania, Sweden, Slovakia, and Slovenia, see European Commission, 2018).

[6]Other Clean Air Plan measures typically aim at enhancing the use of public transportation, bicycles or electric powered vehicles and are much less specific.

[7]As a result of court decisions, as of 2019 access to certain Low Emission Zones or other specific city areas (e.g., in Stuttgart) requires a minimum emission standard of Euro 5 by diesel-fueled vehicles.

quality standards within a city area, Low Emission Zone policies are exogenously enforced upon cities either by the responsible state governments or court rulings based on EU air quality legislation.

A Low Emission Zone is a signposted area where entry by vehicles is regulated, usually by prohibiting vehicles with higher emissions from entering the area altogether. Access regulation is based on the six emission standards based on EU legislation. The emission standard of a vehicle is categorized by color-coded windscreen stickers with no sticker for the highest emission level Euro 1 and red, yellow and green stickers for “cleaner” emission standards Euro 2–4 (see Table 2.A.1.2 in Appendix 2.A.1 for details). Typically, Low Emission Zones are introduced in phases. In phase one, only the dirtiest Euro 1 vehicles were banned. Subsequently, the Low Emission Zones became stricter, banning Euro 2 and Euro 3 classes in the second phase and finally allowing only green sticker (Euro 4) vehicles in the third phase. As of 2018, there are 58 Low Emission Zones in Germany with only one being accessible by vehicles displaying a yellow sticker, whereas all remaining Low Emission Zones allow access only to vehicles with a green sticker (see Table 2.A.1.2 in Appendix 2.A.1 for an overview).[8]

We use data on all Low Emission Zones in Germany from the Federal Environment Agency (*Umweltbundesamt, UBA*) on the history of implementation by stage (ban of Euro 1–3 vehicles) as well as the precise geographic coverage of each zone at all stages.[9] Figures 2.1 and 2.2 show the spatial diffusion as well as the number of implemented Clean Air Plans and Low Emission Zones over the period from 2007 to 2018. The first Clean Air Plans were established in 2007, the number increased to more than 80 by 2018. In 2008, eleven Low Emission Zones were established at stage one (only banning Euro 1 vehicles) followed by a gradual increase of new Low Emission Zones across the country. The earliest second stage (banning Euro 1–2) was introduced in 2009, while over the course of 2010 all Low Emission Zones switched at least to the second stage, some already introduced the third stage (ban on Euro 1–3). From 2013 onwards, the third stage dominated. As of 2018, there are 58 active Low Emission Zones in Germany. Whereas in 2018 Clean Air Plans are rather equally distributed across Germany, most Low Emission Zones are located in urban areas in the West or South-West of Germany.

[8] In 2018, the penalty for violation is 80 Euros. The Low Emission Zone policies are enforced by the police and by local public order authorities. Two-wheeled vehicles, vintage cars, police, fire brigade and emergency vehicles and farm machinery are exempt from the scheme.

[9] We use open source polygons of Low Emission Zones in German cities from OpenStreetMap.org. As an example, Figure 2.A.1.3 shows the high congruency with official documentation for the largest Low Emission Zone in the Ruhr area.

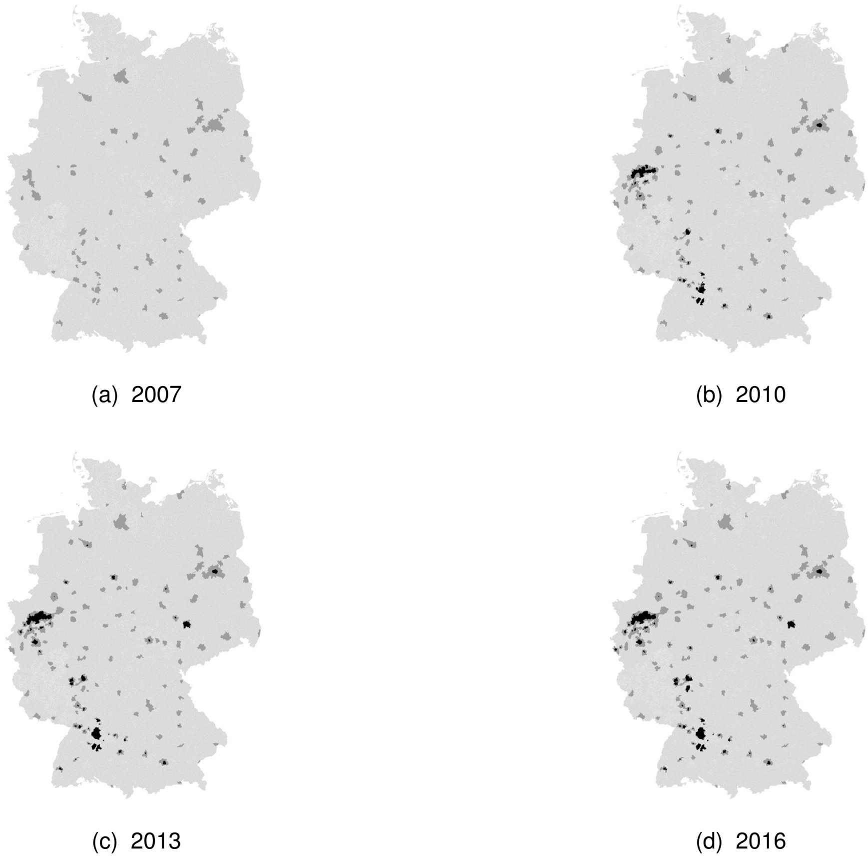


Figure 2.1: Clean Air Plans (grey) and Low Emission Zones (black) over time

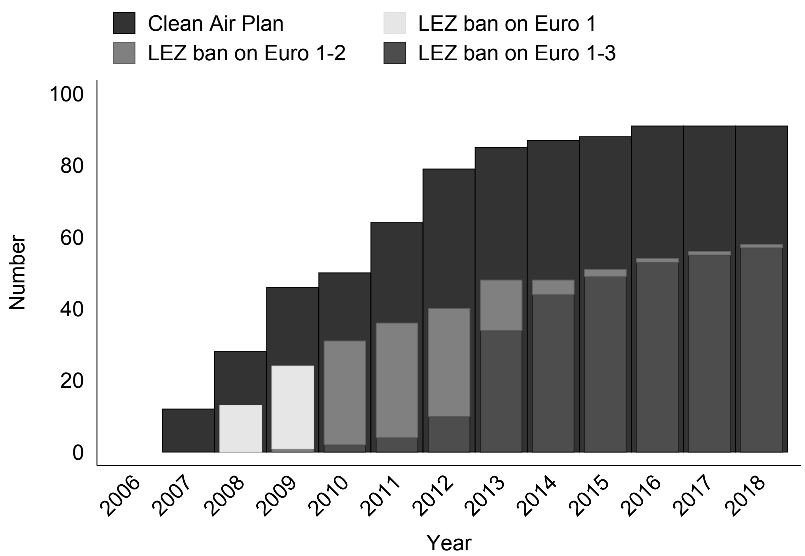


Figure 2.2: Clean Air Plans and Low Emission Zones by emission standard over time

2.2.2 Air pollution: Risks to human health and measurement

The purpose of Low Emission Zones is to improve air quality in urban areas by reducing the emission of harmful air pollutants from traffic.[10] The main target air pollutants emitted from traffic are particulate matter (PM) and nitrogen dioxide (NO2).[11] In the following, we explain how these air pollutants are generated and how they may affect human health.

Particulate matter (PM) measures the concentration of small airborne particles including dust, dirt, soot, smoke and liquid droplets which are emitted to ambient air from a variety of sources. Natural sources are bush fires, dust storms, pollens and sea spray while anthropogenic sources include motor vehicle emissions and industrial processes. Small particulates may enter the lungs, the smallest particles may even enter the blood stream and overcome the blood-brain barrier causing inflammation. We focus on PM10, i.e., the concentration of particles that are smaller than 10 μm in diameter, which has been comprehensively measured since 2000 in Germany.[12] Particulate matter is linked to a number of respiratory and cardiovascular diseases, among others ischemic heart diseases (which may lead to heart attacks), cerebrovascular diseases (e.g. strokes), chronic and acute lower respiratory diseases as well as low birth weight among newborns (Kampa and Castanas, 2008; Block and Calderon-Garciduenas, 2009).[13]

Nitrogen dioxide (NO2) results from burning fossil fuels like coal, oil and gas. In cities, the major source of nitrogen dioxide is motor vehicle exhaust (up to 80 percent, see Environmental Protection Agency, 2016). Nitrogen dioxide contributes to the formation of photochemical smog, which can have significant impacts on human health (Vitousek et al., 1997). Nitrogen oxides are often linked to nose and throat irritation, and increase sensitivity to respiratory infections (Kampa and Castanas, 2008). Exposure to elevated NO2 concentration in ambient air especially causes respiratory problems by inflaming the lining of the lungs.[14] Based on a systematic literature review Schneider et al. (2018) identified

[10] This may be achieved by reducing traffic volume, by decreasing the vehicle fleet's share of high-emission cars or a combination of both. Wolff (2014) shows that Low Emission Zone introductions in German cities encouraged a shift to a less emitting car stock. Additionally, Figure 2.A.1.4 shows that in Germany the vehicle fleet has become substantially cleaner in terms of average PM10 and NO2 emissions since the mid-1990s. In particular, average emissions of trucks decreased by more than 80 percent. NO2 emissions of cars decreased since 2007 but remained rather constant while PM10 emissions further decreased.

[11] These specific air pollutants are usually used as markers for the cocktail of combustion related pollutants emitted by road traffic. They are highly correlated with each other and associated with other combustion products, such as ultrafine particles, nitrous oxide (NO) or benzene (WHO, 2006). In addition, traffic contributes to the emission of greenhouse gases which are harmful to the climate.

[12] The concentration of fine particles smaller than 2.5 μm (PM2.5) has been regulated by the EU only since 2015.

[13] Particulate pollution from any source has negative impacts on health. However, anthropogenic sources, especially those emitted by traffic, like rubber abrasion, brake dust or exhaust emissions are more harmful (WHO, 2006).

[14] Janke (2014) showed that a one percent increase in NO2 lead to roughly 0.1 percent increase in emergency respiratory hospitalizations for children.

possible NO2 cause-specific hospital admissions: cardiovascular and respiratory morbidity, hypertension, ischemic heart diseases and low birth weight.

Data on air pollution comes from the air pollution monitoring system of the German Federal Environment Agency. We use data on all geo-coded monitors measuring the concentration of particulate matter (PM10) or nitrogen dioxide (NO2) between 2006 and 2016. The main variables of interest are the yearly averages of pollutants as well as yearly number of monitor-specific limit-exceedances and violations according to the EU air quality standards (see Table 2.A.1.1).

Overall, we have 4,290 and 5,237 monitor-by-year observations for PM10 and NO2 respectively. Panels A of Tables 2.1 and 2.2 show that, on average, the yearly mean levels of PM10 and NO2 pollution are well below the limit values of 40 micrograms per cubic meter ($\mu g/m^3$). The yearly mean of PM10 is 22 $\mu g/m^3$ and 31 $\mu g/m^3$ for NO2. However, there is sizable variation between monitors within years as well as within monitors by year which leads to violations of the EU air quality standards by exceeding the maximum number of days or hours with higher concentrations. For example, in about eight percent of monitor-year observations there are more than 35 days with a daily mean PM10 concentration of 50 $\mu g/m^3$ and 30 percent of observations exceed the annual mean NO2 limit.[15]

Combined with the data on Low Emission Zones, we are able to assign whether a monitor is located inside or outside of a Low Emission Zone area and, if outside, to compute the distance to the closest Low Emission Zone boundary.[16] Panels B of Tables 2.1 and 2.2 show that between 2006 and 2016, more than half of pollution monitor observations are covered by an active Clean Air Plan. The share of observations covered by a Low Emission Zone banning at least Euro 1 vehicles (red sticker) is 14–18 percent, the share is 9–12 percent for Low Emission Zones banning at least Euro 1–2 (yellow sticker) and 7–10 percent for Low Emission Zones banning Euro 1–3 (green sticker).

Further control variables for the sample of pollution monitors are shown in Panels C and D of Tables 2.1 and 2.2. Since weather conditions are important environmental confounders we further supplement our dataset with a rich set of weather controls. The data are provided by the German Meteorological Service (*Deutscher Wetterdienst*) and contain information on temperature, precipitation and wind speed. We retrieve the yearly averages at the closest weather station for each pollution monitor to control for confounding effects. Finally, we control for a number of population characteristics provided by the Federal Statistical Office at the level of the municipality of the pollution monitor location.

[15] Figure 2.A.1.5 shows how pollution levels and violations evolved over time.

[16] Figure 2.A.1.6 a shows that the location of pollution monitors across Germany largely reflects more densely populated urban areas, which are also typically covered by Clean Air Plans and Low Emission Zones (see Figure 2.1).

Table 2.1: Descriptive statistics: Sample of PM10 monitors (2006–2016)

	Mean	SD	Min	Max	N
A. Pollution outcomes					
Yearly mean PM10 ($\mu g/m^3$)	21.94	5.68	7.00	55.00	4290
Yearly days PM10 > 50 $\mu g/m^3$	15.41	14.18	0.00	175.00	4290
Violation (Yearly mean PM10 > 40 $\mu g/m^3$)	0.00	0.06	0	1	4290
Violation (Days PM10 > 50 $\mu g/m^3$)	0.08	0.28	0	1	4290
B. Treatment characteristics					
In active Clean Air Plan	0.54	0.49	0	1	4290
In LEZ ban on Euro 1	0.14	0.33	0	1	4290
In LEZ ban on Euro 1-2	0.09	0.28	0	1	4290
In LEZ ban on Euro 1-3	0.07	0.25	0	1	4290
C. Weather characteristics					
Mean temperature ($^\circ$C)	9.70	1.44	2.75	12.78	4290
Mean precipitation (mm/m^2)	2.05	0.61	0.54	5.82	4290
Mean Wind speed (m/ss)	3.46	0.98	1.66	11.19	4290
D. Municipality characteristics					
Inhabitants/1000	151.06	453.30	0.04	3574.83	4290
Employed/1000	65.76	182.65	0.00	1367.68	4290
Share male < 30 years	0.32	0.03	0.23	0.41	4290
Share male 30 - 64 years	0.50	0.02	0.43	0.55	4290
Share male > 64 years	0.18	0.02	0.13	0.27	4290
Share female < 30 years	0.29	0.03	0.20	0.39	4290
Share female 30 - 64 years	0.47	0.02	0.41	0.52	4290
Share female > 64 years	0.24	0.03	0.17	0.34	4290

Notes: This table displays the descriptive statistics for the most important variables. The data underlying the statistics in Panel C are measured at the nearest measuring station to the pollution monitor. Panel D is based on the municipality a monitor is located at.

Table 2.2: Descriptive statistics: Sample of NO2 monitors (2006–2016)

	Mean	SD	Min	Max	N
A. Pollution outcomes					
Yearly mean NO2 ($\mu g/m^3$)	30.86	21.98	0.00	121.35	5237
Yearly hours NO2 > 200 $\mu g/m^3$	2.07	24.73	0.00	853.00	4365
Violation (Yearly mean NO2 > 40) $\mu g/m^3$)	0.30	0.46	0	1	5237
Violation (Hours NO2 > 200 $\mu g/m^3$)	0.02	0.13	0	1	4365
B. Treatment characteristics					
In active Clean Air Plan	0.59	0.49	0	1	5237
In LEZ ban on Euro 1	0.18	0.37	0	1	5237
In LEZ ban on Euro 1-2	0.12	0.31	0	1	5237
In LEZ ban on Euro 1-3	0.10	0.29	0	1	5237
C. Weather characteristics					
Mean temperature ($^\circ$C)	9.71	1.47	0.48	12.78	5237
Mean precipitation (mm/m^2)	2.09	0.63	0.54	7.52	5237
Mean Wind speed (m/ss)	3.47	1.01	1.44	11.25	5237
D. Municipality characteristics					
Inhabitants/1000	158.24	443.57	0.04	3574.83	5237
Employed/1000	69.09	179.61	0.00	1367.68	5237
Share male < 30 years	0.32	0.03	0.23	0.41	5237
Share male 30 - 64 years	0.50	0.02	0.43	0.55	5237
Share male > 64 years	0.18	0.02	0.13	0.27	5237
Share female < 30 years	0.29	0.03	0.20	0.39	5237
Share female 30 - 64 years	0.47	0.02	0.41	0.52	5237
Share female > 64 years	0.24	0.03	0.17	0.34	5237

Notes: This table displays the descriptive statistics for the most important variables. The data underlying the statistics in Panel C are measured at the nearest measuring station to the pollution monitor. Panel D is based on the municipality a monitor is located at.

2.2.3 Diagnoses from the universe of German hospitals

For our analysis of health effects from Low Emission Zones we use a panel dataset of the universe of hospitals in Germany reporting the annual number of detailed diagnoses for in-patient cases.[17] German hospitals are obliged by law to publish structured quality reports since 2006, every second year until 2012 and annually from 2012 onwards. The structure and content of these reports are specified legally and misreporting leads to financial penalties. The reported data provide information on structure and performance of a hospital at the hospital department level. The quality reporting was implemented to demonstrate hospitals' performance in a transparent manner to enable a well-informed choice of hospitals by patients, as well as to guide and support referring physicians and sickness funds.[18]

Hospital quality report data comprise hospital characteristics like the number of beds and ownership structure, as well as the yearly number of in-patient cases and diagnoses based on the full International Statistical Classification of Diseases and Related Health Problems (ICD-10). Given that the data's intention is to increase transparency, every hospital is non-anonymously identified, allowing us to assign the treatment of coverage by a Low Emission Zone at the exact address location.[19] The full dataset includes more than 2,000 hospitals over the period from 2006 to 2016 (see Figure 2.A.2.1 in Appendix 2.A.2). We exclude hospitals that do not meet the criteria of hospitals of primary care in Germany (*Krankenhäuser der Regelversorgung*), i.e., having a unit for surgery and internal medicine (Ethikrat, 2016). Hence, we focus on **general hospitals** and exclude specialized hospitals like hospices, wellness clinics, rehabilitation centers, sanatoriums etc., resulting in a sample of around 1,100 hospitals per year and 8,828 hospital-year observations. This reduces measurement error because the excluded hospitals perform an over-proportional amount of planned treatments where spatial proximity is less crucial and often do not treat air pollution related diseases (Klauber et al., 2015).[20]

Panel A of Table 2.3 shows substantial variation in the characteristics of general hospitals. The mean number of beds ranges from only four to 2,917, revealing that the definition of a hospital is independent of its size but rather a legal concept based on permanent availability and equipment. In-patients per year range from 77 to 198,452 with a mean

[17] Admissions to a hospital are usually due to more severe health issues. Therefore, hospitalization data does not cover milder medical conditions which are reflected in doctor visits (if at all). In-patient cases are even more severe because hospitals are obliged to justify that an outpatient treatment is not sufficient. Otherwise, they jeopardize the full reimbursement by health insurances. However, hospital discharge rates in Germany are relatively high, also due to the fact that Germany is among the countries with the highest hospital density (Kumar and Schoenstein, 2013).

[18] See Appendix 2.A.2 for a detailed description of the data.

[19] We use the HERE navigation API to convert full addresses into geocodes.

[20] The robustness checks includes an analysis of the specialized hospitals as well as specifications where we only include hospitals located in cities that ever adopted a Clean Air Plan to make the control group more comparable to the treatment group.

of 15,669. Non-profit and public general hospitals account for 43 and 40 percent in our dataset. About 17 percent of the general hospitals in our dataset are private. However, private general hospitals in Germany are obliged to provide the same health services to the same conditions as non-private.[21]

The total number of diagnoses according to the ICD-10 classification (indicated in brackets) are shown in Panel B of Table 2.3. The average number of annual diagnoses of diseases is 10,506. We mainly focus on the overall number of diagnosed diseases of the circulatory system, making up 22 percent of all diseases, and the respiratory system (about nine percent), which are also broken down to more detailed ICD-10 subcategories. In addition, we will look at low birth weight as an outcome (Gehrsitz, 2017) as well as stress-related diagnoses and to overall number of injuries potentially reflecting changes in the number of traffic accidents due to potentially lower traffic volume caused by Low Emission Zone restrictions of vehicle entry to the area.

Hospital catchment areas are assigned based on hospitals' locations since the hospital quality report data does not provide information on the residence of patients. While there is a free choice of hospitals in Germany, a strong correlation exists between hospital location and patients' residences (Friedrich and Beivers, 2008). In general, individuals prefer hospitals close to their residential address (Klauber et al., 2015). Furthermore, resident doctors are legally obliged to refer patients to one of the two closest hospitals based on their residence. Knowing the location of the hospital is even more advantageous when analyzing more severe emergency cases in which admission is based on the patient's current position which is not necessarily equal to the place of residence (Klauber et al., 2008). In 2016, 45 percent of hospital admissions were emergency cases (Statistisches Bundesamt, 2017a).[22] According to the directive for ambulance transport (*Krankentransport-Richtlinie*), emergencies should be transported directly to the nearest hospital.

For our main analysis, we use the Open Source Routing Machine (OSRM) with the OpenStreetMap road network of 2016 to define mutually exclusive hospital catchment areas based on driving time. This means that for every hospital in our dataset, we create adjacent polygons around the hospital location corresponding to regions comprising all points that have a shorter driving time to the hospital than to any other hospital in the surrounding. These regions are the catchment areas. Hence, each point on a border between two catchment areas has the exact same driving time to the two corresponding hospital locations.[23]

[21] Three types of hospital ownership are defined by German Law: public, owned by the state, a federal state or a municipality; non-profit, Owned by non-profit organizations like the Red Cross or institutions of the churches and private, primarily aim at making a profit by individuals or legal entities (Wissenschaftliche Dienste, 2014).

[22] The statistics do not allow distinguishing self-referral from referral by emergency services

[23] Mutually-exclusive driving time polygons are a well established technique to define hospital catchment areas (McLafferty, 2003) and has been validated for such approaches (Schuurman et al., 2006). Figures 2.A.2.2 and 2.A.2.3 in Appendix 2.A.2 shows the location of hospitals and their catchment areas across Germany and zoomed in for the city

As these catchment areas do not perfectly map into administrative geographic areas we do not have information on their population. In a robustness check, we account for heterogeneity in population size by weighting with approximated population density from high resolution satellite data, which is only available for two years over the period of investigation.

Hospitals' treatment by Low Emission Zones is assigned by overlaying the Low Emission Zone areas with the hospital locations and catchment areas. Panel C of Table 2.3 shows that one third of hospital-year observations are located in a municipality with an active Clean Air Plan and ten percent of the observations in Low Emission Zones that ban at least Euro 1 vehicles, seven percent banning at least Euro 1–2 and six percent banning Euro 1–3. Further, we calculate the proportion of a hospital catchment area that is covered by an active Low Emission Zone. At the extensive margin, 16 percent of all general hospital observations have a catchment area that is at least partly covered by an active Low Emission Zone. The overall share of catchment areas that is covered by Low Emission Zones is six percent. In section 2.3.4 we provide a series of robustness checks where we use different treatment specifications to account for measurement error.

Figure 2.A.2.4 in Appendix 2.A.2 reveals a constant increase of hospitals which are located in Low Emission Zones over time. Whereas in 2006 no hospital was located in a Low Emission Zone this share increased to 13 percent in 2016. Hospitals whose catchment areas overlap with Low Emission Zones account for almost 22 percent of all hospitals. This trend is partly driven by a trend of urbanization of hospital supply (Klauber et al., 2015).

Bonn in West Germany.

Table 2.3: Descriptive statistics of hospital characteristics

	Mean	(SD)	min	max	N
A. Hospital characteristics					
Non-profit	0.43	0.50	0	1	8828
Public	0.40	0.49	0	1	8828
Private	0.17	0.38	0	1	8828
Number of Beds	375.49	312.82	4	2917	8828
Base rate in €	2990.23	260.91	871	14238	8828
Inpatients	15669.04	14263.88	77	198452	8828
Catchment area in km^2	503.51	559.65	0	4671	8828
Population in catchment area	75859.64	54525.85	282	447094	8820
B. Diagnoses					
All diseases (A00-N99)	10506.28	10257.91	32	155406	8828
Diseases of the circulatory system (I00-I99)	2294.15	2579.06	0	55735	8828
Hypertension (I10-I15)	258.84	398.95	0	18855	8828
Ischemic heart diseases (I20-I25)	565.18	867.48	0	17668	8828
Cerebrovascular disease (I60-I69)	277.00	420.80	0	6118	8828
Diseases of the respiratory system (J00-J99)	944.61	989.36	0	15512	8828
Chronic lower respiratory diseases (J40-J47)	203.95	221.67	0	3812	8828
Acute lower respiratory diseases (J20-J22)	103.91	122.82	0	1392	8828
Low birth Weight (P07) [t+1]	46.09	104.92	0	1840	7507
Stress (F40-F48)	74.71	141.68	0	2614	8828
Injuries (S00-S99)	1185.00	1119.20	0	19174	8828
C. Treatment characteristics					
In active Clean Air Plan	0.34	0.47	0.00	1.00	8828
In LEZ ban on Euro 1	0.10	0.30	0.00	1.00	8828
In LEZ ban on Euro 1-2	0.07	0.26	0.00	1.00	8828
In LEZ ban on Euro 1-3	0.06	0.23	0.00	1.00	8828
Catchment areas covered by LEZ	0.16	0.37	0.00	1.00	8828
Overall share of catchment area covered by LEZ	0.06	0.20	0.00	1.00	8828
Overall share of population covered by LEZ	0.07	0.22	0.00	1.00	8828
D. Weather characteristics					
Mean temperature (°C)	9.63	1.43	-5.27	12.64	8828
Mean precipitation in mm/m^2	2.05	0.58	0.80	5.89	8828
Mean Wind speed (m/ss)	3.42	0.98	1.18	11.19	8828
E. Municipality characteristics					
Inhabitants/1000	263.36	634.02	0.40	3574.83	8828
Employed/1000	113.82	249.37	0.00	1367.68	8828
Share male < 30 years	0.32	0.03	0.23	0.41	8828
Share male 30 - 64 years	0.50	0.02	0.43	0.55	8828
Share male > 64 years	0.18	0.02	0.13	0.27	8828
Share female < 30 years	0.29	0.03	0.20	0.39	8828
Share female 30 - 64 years	0.47	0.02	0.41	0.53	8828
Share female > 64 years	0.23	0.03	0.16	0.34	8828

Notes:This table displays the descriptive statistics for the most important variables. The data underlying the statistics in Panel D are measured at the nearest measuring station to the hospital. Panel E is based on the municipality a hospital is located in.

2.3 Empirical Analysis

2.3.1 Regression Model

Our aim is to estimate the causal impact of the introduction of a Low Emission Zone (LEZ) on population health via improvements in air quality. The staggered introduction of Low Emission Zones across cities in Germany motivates a difference-in-differences estimation strategy with the following empirical model, which we apply to both the sample of air pollution monitors and the sample of hospitals in Germany over the period 2006–2016. The basic model reads:

$$y_{ict} = \alpha + \beta \; LEZ_{it} + \boldsymbol{X'_{ict}}\gamma + \delta_i + \delta_{ts(c)} + \varepsilon_{ict}, \tag{2.1}$$

where y_{ict} indicates the outcome – a measure of air pollution or the number of diagnoses – in year t measured at observation unit i – a pollution monitor or a hospital – located in city c. The main variable of interest is LEZ_{it} and captures the treatment of unit i in year t by a Low Emission Zone, which differs depending on the sample. For the sample of air pollution monitors, LEZ_{it} is simply a binary indicator with a value of one for monitor i being located within the boundaries of an active Low Emission Zone at any strictness level in year t and zero otherwise.[24] For the sample of hospitals, we equate LEZ_{it} with the share of hospital i's catchment area covered by a Low Emission Zone, ranging between zero and one.

The vector $\boldsymbol{X_{ict}}$ controls for a number of time-varying characteristics at the level of monitors and hospitals as well as for city population characteristics. In both samples, we include the set of weather controls measured at the closest weather monitor to the pollution monitor or hospital respectively. Further, we include population size, employment as well as the city population's composition by age groups and gender (see Tables 2.1–2.3 for details). For the sample of hospitals, we further control for time-varying hospital characteristics, the number of hospital beds, ownership and the baserate.[25] Finally, unit fixed effects δ_i capture any time-invariant monitor or hospital characteristics while state-year fixed effects $\delta_{ts(c)}$ control for any time-specific effects that are uniform across all observation units within a state s. To capture urbanization processes we also include city-specific linear time trends. The error term ε_{ict} is clustered at the county level.[26]

[24] In the Appendix, we show that the reductions in pollution are rather mixed across Low Emission Zone strictness levels. Further, most Low Emission Zones were introduced on January 1. If not, we multiply LEZ_{it} by 0.5 if the Low Emission Zone was established not later than June 30 in the introduction year t and set LEZ_{it} to zero if the Low Emission Zone was introduced later than June 30.

[25] The number of beds per hospital are determined annually at the regional level by hospitals, insurance associations and regional administrations to ensure sufficient supply based on population. The baserate reflects the historic cost level and determine hospital specific reimbursement prices.

[26] In Germany, larger cities are identical to a county (*Kreisfreie Stadt*), while more rural counties (*Landkreise*) comprise multiple smaller cities.

In order to capture dynamic effects of Low Emission Zone introductions, we conduct event studies where we test whether Low Emission Zone effects differ over the post-treatment periods. In addition, this allows to test whether the identifying assumption of common pre-trends is violated. The introduction of a Low Emission Zone should not have any impact in pre-treatment periods. The extended model is:

$$y_{ict} = \alpha + \sum_{k=-4,k\neq-1}^{+5} \beta^k \; LEZ_{ik} + \boldsymbol{X'_{ict}}\gamma + \delta_i + \delta_{ts(c)} + \varepsilon_{ict}, \tag{2.2}$$

where the dummy variables LEZ_{ik} indicate yearly leads and lags of up to four years before and five years after the enactment of a Low Emission Zone. The reference category is $k = -1$, hence the post treatment effects are relative to the year immediately before the policy change and are interpreted as the effect of Low Emission Zones k periods before or after their introduction. We use the same controls as before.[27]

2.3.2 The impact of Low Emission Zones on air quality

In a first step, we document how the implementation of Low Emission Zones affects local air pollution by regulating the entry of vehicles based on their emission exhaust. Table 2.4 shows the main results for the effect of introducing a Low Emission Zone on annual average levels, limit exceedances and violations for PM10 and NO2. Each entry represents an estimate for β according to equation (2.1) from a separate regression of the respective outcome on the Low Emission Zone indicator, i.e., for a monitor being located within the boundaries of an active Low Emission Zone.

The results in Panel A of Table 2.4 show a negative impact on pollution levels for both PM10 and NO2 concentrations in all three specifications where we start with a fixed effect regression and gradually add time-variant control variables. Controlling for weather characteristics does not change the estimates. By adding additional controls for municipality characteristics effect sizes for most coefficients slightly decrease in absolute terms by capturing different changes in demographic compositions between areas. This is why we prefer the specification in columns (3) and (6) in the following analysis. The introduction of a ban of at least Euro 1 emission classes decreases PM10 by 1.3 $\mu g/m^3$ or six percent of the mean. The average NO2 levels are reduced by 1.6 $\mu g/m^3$ or five percent of the mean. Both effects are statistically significant at the one percent level.

In Panel B, we show results on outcomes related to limit exceedances according to the air quality standards. Introducing a Low Emission Zone reduces the annual number of days with PM10 levels above the regulatory threshold of 50 $\mu g/m^3$ by 7.7 days or almost 50 per-

[27] We bin up event dummies at the endpoints of the event window (i.e., $k = -4$ and $k = 5$). Hence, these dummies account for Low Emission Zone effects four or more years before and five or more years after the introduction

cent of the mean. In Panel C, we do not find any effect on the incidence of the yearly PM10 mean being above 40 $\mu g/m^3$, which is an extremely rare event to begin with (see Table 2.1). Although negative, we do not find statistically significant effects on limit exceedances for yearly hours of NO2 > 200 $\mu g/m^3$. Again, the incidence of violating this threshold is relatively rare (Table 2.2). However, we do find a significant decrease of yearly mean NO2 levels above 40 $\mu g/m^3$ of about four percentage points, which corresponds to a sizable reduction of about 25 percent compared to the mean. Hence, the policy of introducing a Low Emission Zone appears to be very effective in significantly decreasing local air pollution and reducing the incidence of air quality standard violations. Introducing Low Emission Zones effectively reduces the incidence of short-time spikes in PM10 pollution and at the same time reduces the longer-term annual mean concentration of NO2.

These findings are based on the straightforward specifications of equation (2.1), where we exploit the treatment of any Low Emission Zone irrespective of the strictness levels in terms of the emission exhaust classification.[28] While Low Emission Zone introductions typically begin with banning the dirtiest Euro 1 vehicles from entering the inner-city areas, essentially all Low Emission Zones by now ban Euro 1–3 vehicles. In Table 2.A.1.4 we show results for interacting the Low Emission Zone treatment with different strictness levels, i.e., banning Euro 2 and Euro 3 additionally. It turns out that all strictness levels contribute to the average effects for most pollution outcomes shown in Table 2.4. These results also reflect the general improvement of emissions from vehicles and an upgrade of the vehicle fleet towards lower emission cars in cities with Low Emission Zones (Wolff, 2014) since more restrictive Low Emission Zones have been implemented later in time (see Figure 2.A.1.4 in Appendix 2.A.1). The spatial precision of our dataset allows us to analyze the effect of a Low Emission Zone on pollution in its surroundings. Table 2.A.1.5 in Appendix 2.A.1 shows that air quality in close proximity to a Low Emission Zone (within a radius of 10 km) is not affected while do some smaller increases for pollution monitors located at a distance of 10–20 km from a Low Emission Zone.

In Figure 2.3 we present results for the event study specification of equation (2.2). Focusing on those pollution outcomes with statistically significant effects as shown in Table 2.4, we use the presence of an active Low Emission Zone at the location of a pollution monitor as treatment independent of its strictness with the reference period $k = -1$, the year before a Low Emission Zone became effective. The event study results do not reveal any pre-trends that could bias our results. Corresponding to the difference-in-differences estimates, we find that air pollution levels as well as the incidence of violating regulatory thresholds

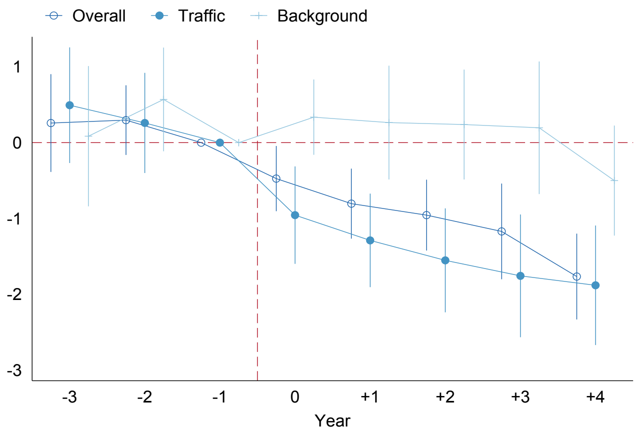
[28] In Table 2.A.1.3 we present results on the effects of introducing a Clean Air Plan, typically preceding Low Emission Zones by a few years, interacted with the introduction of a Low Emission Zone. We find that Clean Air Plans indeed have a negative effect on air pollution but that this is mainly driven by Low Emission Zone introductions. However, we refrain from putting too much emphasis on these findings since Clean Air Plans are very heterogeneous measures with unclear spatial extent.

for air quality are significantly reduced right after after the introduction of a Low Emission Zone. With the exception of the yearly mean of NO2 being above 40 $\mu g/m^3$ the effects become stronger over time. This could be due to the fact that Low Emission Zones have become stricter over time (see also Table 2.A.1.4 in Appendix 2.A.1). In addition, Figure 2.3 shows results for splitting the sample of pollution monitors by whether they are designated as traffic or background monitors. As expected, the reductions in air pollution are strongest for traffic monitors.

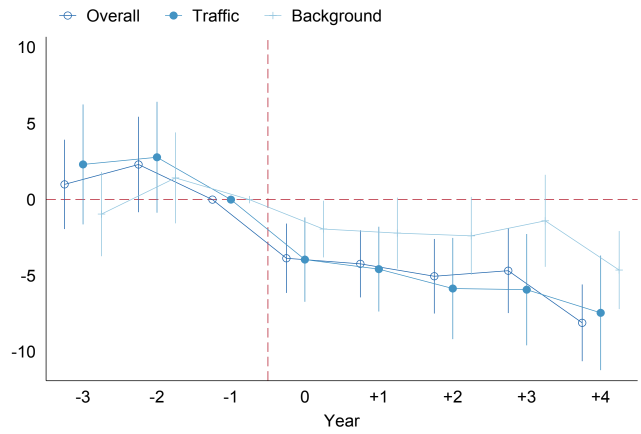
Table 2.4: The effect of Low Emission Zones on air pollution

	PM10			NO2		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Pollution levels	Yearly mean PM10 ($\mu g/m^3$)			Yearly mean NO2 ($\mu g/m^3$)		
In LEZ	-1.461***	-1.450***	-1.273***	-1.829***	-1.823***	-1.581***
	(0.188)	(0.188)	(0.204)	(0.454)	(0.455)	(0.461)
Adj. R^2	0.93	0.93	0.93	0.74	0.74	0.74
N	4290	4290	4290	5237	5237	5237
B. Limit exceedances	Yearly days PM10 > 50 ($\mu g/m^3$)			Yearly hours NO2 > 200 ($\mu g/m^3$)		
In LEZ	-7.068***	-7.068***	-6.580***	-7.071	-7.061	-5.572
	(0.916)	(0.918)	(0.972)	(4.783)	(4.785)	(3.883)
Adj. R^2	0.81	0.81	0.82	0.48	0.48	0.50
N	4290	4290	4290	4365	4365	4365
C. Violations	Yearly mean PM10 > 40 ($\mu g/m^3$)			Yearly mean NO2 > 40 ($\mu g/m^3$)		
In LEZ	-0.003	-0.003	-0.000	-0.050**	-0.051**	-0.043**
	(0.007)	(0.007)	(0.006)	(0.021)	(0.021)	(0.022)
Adj. R^2	0.16	0.17	0.17	0.86	0.86	0.86
N	4290	4290	4290	5237	5237	5237
Controls:						
Station FE	Yes	Yes	Yes	Yes	Yes	Yes
State $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather characteristics	No	Yes	Yes	No	Yes	Yes
Municipality characteristics	No	No	Yes	No	No	Yes

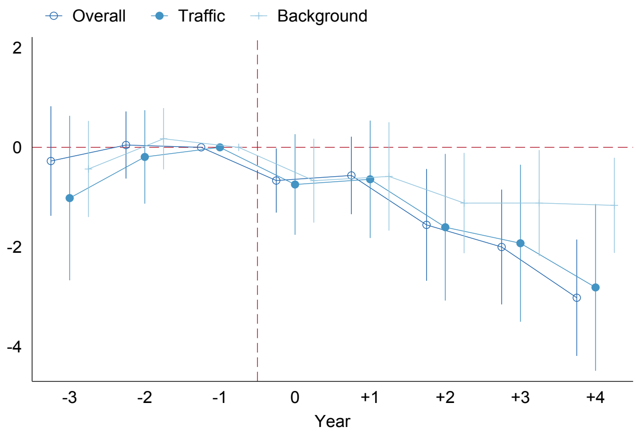
Notes: This table displays the results for the effect of Low Emission Zones on air pollution. Each coefficient is the result of a separate regression of pollution levels listed on the left on an indicator variable for locates in an active Low Emission Zone, while controlling for monitor and year fixed effects as well federal state time trends, weather characteristics characteristics (mean temperature, precipitation and wind speed) and municipality characteristics (population, work force, age structure (share man(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max). Standard errors are clustered at county level and displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.



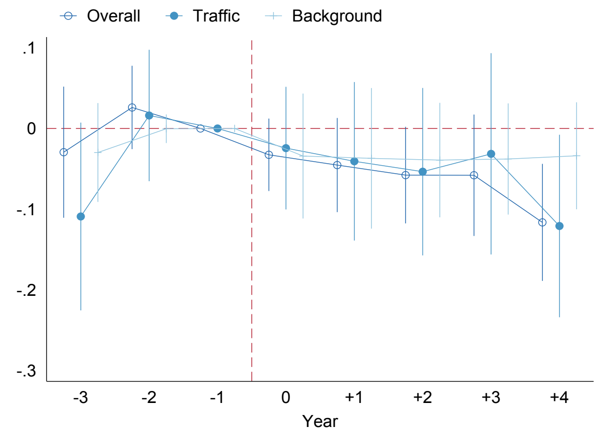
(a) Mean PM10 $\mu g/m^3$



(b) Days PM10 > 50 $\mu g/m^3$



(c) Mean NO2 $\mu g/m^3$



(d) Mean NO2 > 40 $\mu g/m^3$

Figure 2.3: The effect of Low Emission Zones on air pollution (event study)

Notes: This figure displays an event studies for the effect of Low Emission Zones on air pollution. The reference period is $k = -1$. Each coefficient is the result of a separate interactions of dummy variables counting the years before and after the introduction of a Low Emission Zone and an indicator variable showing if a monitor is located inside or outside of a Low Emission Zone, while controlling for monitor and year fixed effects as well as weather characteristics (mean temperature, precipitation and wind speed) and municipality characteristics (population, workforce, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)). Standard errors are clustered at county level.

2.3.3 Health effects of Low Emission Zones

The results presented so far have shown that the introduction of a Low Emission Zone in an inner-city area significantly reduces air pollution and violations of EU air quality standards mainly inside the Low Emission Zone areas. While the EU air quality standards directly target local air pollution one key policy motivation for regulating entry of vehicles to inner-city areas is to improve population health and well-being. After having documented that Low Emission Zones effectively reduce air pollution, we will now turn to the question whether improvements in air quality induced by Low Emission Zones translate into improvements for human health.

In this section, we present the estimation results for the impact of Low Emission Zones on the number of diagnoses per hospital (in logs) as a proxy for general health. Given that the introduction of Low Emission Zones reduced several pollutants at the same time, we are not able to disentangle the effects on diagnoses by pollutant but will focus on hospital diagnoses that are related to PM10 and NO2 (see Section 2.2). Hence, estimates of the β coefficient measure the total effect of a Low Emission Zone introduction on hospital diagnoses. Therefore, results are reduced-form effects. A higher share of a hospital's catchment area covered by a Low Emission Zone lowers potential exposure to air pollution of people living or working in the catchment area. In addition, β captures the direct physiological impact of air pollution on the human body but may also be partly driven by reductions in traffic noise as well as behavioral responses to air pollution, such as changes in exercise habits or internal migration.

Table 2.5 reports the main results for the Low Emission Zone effect on hospital diagnoses. Each cell in this table represents an estimate for β from a separate regression of the outcome listed in the left column on the share of the hospital's catchment area covered by a Low Emission Zone based on driving time and the controls listed at the bottom. We look at the total number of all diseases and then separately at diseases of the circulatory and the respiratory system as well as subgroups thereof. In addition, we use the incidence of low birth weight as an outcome.

We begin with a bivariate fixed effect regression and gradually add control variables. Including hospital fixed effects is particularly important because observable characteristics vary considerably between areas with and without Low Emission Zones since Low Emission Zones are primarily located in densely populated urban areas. By using fixed effects we control for time-invariant structural differences. In addition, we control for a number of time-variant hospital and municipality characteristics and eventually include linear municipality-specific time trends to capture the effects of changing population characteristics over time.

Almost all point estimates in columns (1) to (5) of Table 2.5 are negative irrespective of additional control variables, indicating that the introduction of a Low Emission Zone potentially has a beneficial impact on population health. In columns (2) and (3), once we control for weather and municipality characteristics, coefficient estimates are only slightly affected. Controlling for linear municipality-specific time trends in column (5), capturing differential effects from urbanization, yields to larger point estimates in absolute terms. After including hospital controls in column (5), in particular hospital capacity proxied by the number of beds, most point estimates become larger and more significant. This is not surprising, given that hospitals treated by a Low Emission Zone tend to be located in growing urbanized areas, which increases the number of diagnoses simply because the number of potential patients in the area increases.

Based on the results in column (5), an increase in hospital's catchment area covered by a Low Emission Zone by one standard deviation (corresponding to 20 percentage points, see Table 2.3) reduces the total number of diagnosed diseases by about 1.4 percent, the estimate being statistically significant at the 0.1 level.[29] Focusing on diagnoses that are closely related to air pollution, we find an effect for diseases of the circulatory system by 2.9 percent, or 67 cases at the mean. Among this broad category of diseases, the corresponding effects are between 3.1 and 5.0 percent. The point estimates suggest that a Low Emission Zone has the largest impact on circulatory diseases like ischemic heart diseases and cerebrovascular diseases, implying a reduction of diagnoses between 29 and 11 cases per year at the respective means. We do find a statistically significant effects for the aggregate category of respiratory diseases in general as well which is 2.16 percent. Chronic diseases of the lower respiratory system are significantly reduced by more than four percent (or 8 cases at the mean) for a one standard deviation increase in the Low Emission Zone coverage of a hospital's catchment area while there is a 3.4 percent reduction for acute lower respiratory diseases.[30] We do not find a statistically significant impact on low birth weight.[31]

In Table 2.6 we focus on circulatory and respiratory diagnoses that can be described as medical emergencies or non-emergencies. The selection follows Schreyögg et al. (2014) who identify the top 25 of medical emergency and non-emergency hospital diagnoses between 2007 and 2012, based on diagnose characteristics such as the time between admission and treatment. Out of these 25, we select all circulatory and respiratory diagnoses. The results are shown in Table 2.5. We find that the effect for non-emergency cases is negative and statistically significant while the estimate for emergency cases is much smaller and insignificant. This is in line with our finding that chronic respiratory diagnoses are are more strongly affected than acute diagnoses, indicating that the introduction of Low Emis-

[29]As the dependent variable is in logs, the estimates can be interpreted as changes in percentages. For example, in column (5) of Table 2.5 an increase in the coverage of a catchment area by one standard deviation ($= 0.20$) translates into an effect size of $-0.072 \times 0.20 = 0.014$, i.e., 1.4 percent.

[30]Stronger effect sizes for impacts on the circulatory compared to the respiratory system are in line with findings of a meta study which summarized findings on health effects for traffic related pollutants (Hoek et al., 2013).

[31]We do not observe an effect on stress related diagnoses or injuries, thus not indicating additional health channels but air pollution (Table 2.A.5.2).

sion Zones mainly benefited individuals with a bad health condition who have to be admitted to the hospital less often because of their disease.

In Figure 2.4 we present the results for the log number of respiratory and circulatory diagnoses in an event study framework.[32] The findings for circulatory diseases in Panel (a) indicate that the effects started to appear already in the first year after the introduction and tend to decrease over time. However, focusing on the sample of Low Emission Zones that were introduced until 2011, we find that the effects are more sustainable over the following years, while statistical significance becomes less pronounced. Panel (b) shows the event study for respiratory diseases. In the immediate year after the introduction the incidence of respiratory diseases decreases. However, this effect is not sustainable for the overall sample and also less sustainable for Low Emission Zones introduced earlier. This could be due to the fact that since 2012 the majority of the car fleet in Germany reached at least the Euro 4 emission class for the first time, qualifying cars for a green sticker (Kraftfahrt-Bundesamt, 2018). This means that the introduction of Low Emission Zones becomes a less restrictive policy over time as more cars already fulfill the requirements for entering Low Emission Zones, which reduces the potential impact on vehicle emissions.

[32] We show event studies for every diagnose in the appendix (see Figure 2.A.3.1).

Table 2.5: The effect of Low Emission Zones on diagnoses in general hospitals

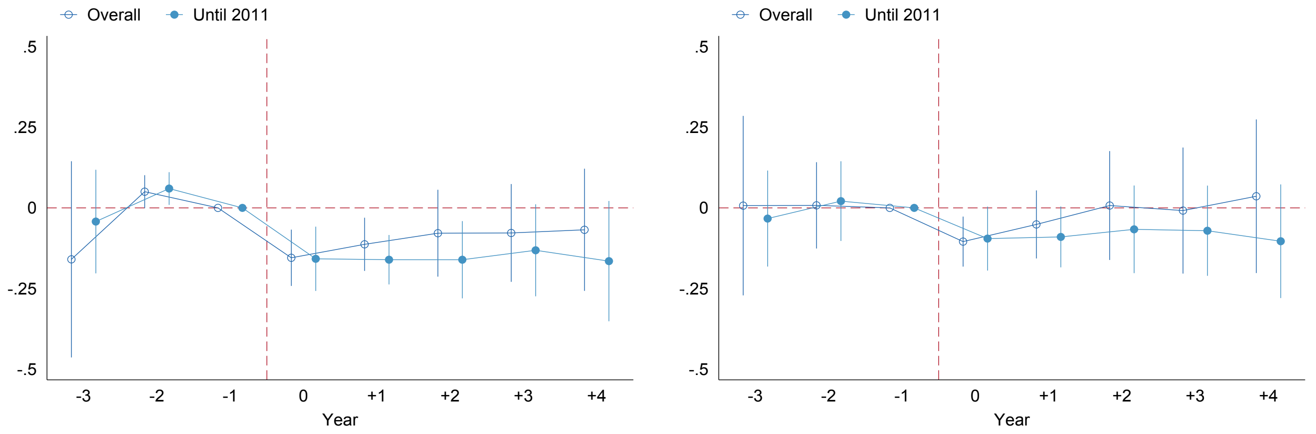
	(1)	(2)	(3)	(4)	(5)
All diseases (A00-N99)	-0.049	-0.050	-0.051	-0.066	-0.072*
	(0.040)	(0.039)	(0.047)	(0.047)	(0.043)
Diseases of the ciculatory system (I00-I99)	-0.105*	-0.105*	-0.109*	-0.144**	-0.145***
	(0.055)	(0.055)	(0.061)	(0.058)	(0.055)
Hypertension (I10-I15)	-0.120*	-0.119*	-0.120	-0.165*	-0.157*
	(0.069)	(0.069)	(0.076)	(0.092)	(0.091)
Ischemic heart diseases (I20-I25)	-0.132	-0.133	-0.174*	-0.258**	-0.263***
	(0.082)	(0.082)	(0.091)	(0.103)	(0.099)
Cerebrovascular disease (I60-I69)	-0.134	-0.133	-0.188*	-0.191*	-0.208**
	(0.099)	(0.099)	(0.103)	(0.106)	(0.102)
Diseases of the respiratory system (J00-J99)	0.009	0.009	-0.007	-0.099	-0.108*
	(0.060)	(0.059)	(0.060)	(0.065)	(0.059)
Acute lower respiratory diseases (J20-J22)	-0.007	-0.006	-0.007	-0.167*	-0.171*
	(0.089)	(0.088)	(0.086)	(0.095)	(0.090)
Chronic lower respiratory diseases (J40-J47)	-0.135**	-0.136**	-0.149**	-0.217***	-0.222***
	(0.063)	(0.063)	(0.067)	(0.079)	(0.077)
Low birth weight (P07) [t+1]	0.027	0.027	0.027	-0.095	-0.091
	(0.077)	(0.077)	(0.082)	(0.111)	(0.112)
N	8828	8828	8828	8828	8828
Controls:					
Hospital FE	Yes	Yes	Yes	Yes	Yes
State $\times$ Year FE	Yes	Yes	Yes	Yes	Yes
Weather characteristics	No	Yes	Yes	Yes	Yes
Municipality characteristics	No	No	Yes	Yes	Yes
Linear municipality time trends	No	No	No	Yes	Yes
Hospital characteristics	No	No	No	No	Yes

Notes:This table displays the results for diagnoses, for general hospitals. The catchment area is calculated by driving time. Each coefficient is the result of a separate regression of diagnose listed on the left on a indicator variable for an active Low Emission Zone (share of catchment area covered by Low Emission Zone), while controlling for hospital and year fixed effects as well as federal state time trends, hospital characteristics (non-profit, public, private, base rate, number of beds, number of beds^2), hospital size (small, medium, large) $\times$ years, municipality characteristics (mean temperature, precipitation and wind speed, population, work force, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)), linear time trends (Municipality $\times$ Years). Standard errors are clustered at county level and displayed in parentheses. Significance levels: $***: p < 0.01, **: p < 0.05, *: p < 0.1.$ [1]) Based on 7516 observations.

Table 2.6: The effect of Low Emission Zones on diagnoses: Emergency vs. non-emergency diagnoses

	(1)	(2)	(3)	(4)	(5)
Main Emergency Cardiovascular and Respiratory Diagnoses	-0.017	-0.017	-0.033	-0.063	-0.064
	(0.065)	(0.065)	(0.071)	(0.085)	(0.083)
Main Non-Emergency Cardiovascular and Respiratory Diagnoses	-0.116**	-0.117**	-0.126*	-0.143**	-0.142**
	(0.058)	(0.058)	(0.065)	(0.060)	(0.057)
N	8828	8828	8828	8828	8828
Controls:					
Hospital FE	Yes	Yes	Yes	Yes	Yes
State $\times$ Year FE	Yes	Yes	Yes	Yes	Yes
Weather characteristics	No	Yes	Yes	Yes	Yes
Municipality characteristics	No	No	Yes	Yes	Yes
Linear municipality time trends	No	No	No	Yes	Yes
Hospital characteristics	No	No	No	No	Yes

Notes:This table displays the results for emergency (I100, J180, I200, I214, J209, I63, I181) and non-emergency diagnoses of the Cardiovascular and respiratory system (I208, I251, I501, I481, I702, I839, J342, J189), for general hospitals. The catchment area is calculated by driving time. Each coefficient is the result of a separate regression of diagnose listed on the left on a indicator variable for an active Low Emission Zone (share of catchment area covered by Low Emission Zone), while controlling for hospital and year fixed effects as well as federal state time trends, hospital characteristics (non-profit, public, private, base rate, number of beds, number of beds^2), hospital size (small, medium, large) $\times$ years, municipality characteristics (mean temperature, precipitation and wind speed, population, work force, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)), linear time trends (Municipality $\times$ Years). Standard errors are clustered at county level and displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$. [1])



(a) Circulatory diseases

(b) Respiratory diseases

Figure 2.4: The effect of Low Emission Zones on diagnoses in general hospitals (event study)

Notes: Figure 2.4a and 2.4b display event studies revealing the impact of $\beta\ shareLEZ_{it}$ on circulatory diseases (I00-I99) and respiratory diseases (J00-J99). The reference period is $k = -1$. Each coefficient is the result of a separate interactions of dummy variables counting the years before and after the introduction of a Low Emission Zone and an indicator variable showing if the share of a hospital catchment area covered by an active Low Emission Zone, while controlling for hospital and year fixed effects as well as federal state time trends, hospital characteristics (non-profit, public, private, baserate, number of beds, number of beds^2), hospital size (small, medium, large) $\times$ years, weather characteristics (mean temperature, precipitation and wind speed) and municipality characteristics (population, workforce, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)) and linear municipality time trends. Standard errors are clustered at county level

2.3.4 Additional Results and Robustness

Definition of hospital sample. Or main results are based on general hospitals, which are are hospitals with a unit for surgery and internal medicine and hence excluded more specialized hospitals. The results for this baseline sample selection are sown in column (1) of Table 2.7, where we standardize diagnoses to mean zero and standard deviation of one to make results comparable across different samples and specifications. Column (2) shows that estimates for the sample of specialized hospitals, which typically do not treat air pollution related diseases, are not affected by the introduction of Low Emission Zones. Column (3) shows estimation results for a sub-sample of general hospitals that are located in cities that are eventually covered by a Clean Air Plan, which typically precedes the implementation of Low Emission Zones. Hence, hospitals in these cities are more comparable to each other. Focusing on this more homogeneous sample of hospitals reveals similar results as in our main specification, indicating that keeping never adopters in the control group does not increase unobserved heterogeneity.

Assignment of hospital catchment areas. Lacking information on the residential locations of patients, an important potential source of measurement error is the assignment of hospital catchment areas and the extent to which they are covered by Low Emission Zones. We perform a series of robustness checks by applying different alternative definitions of catchment areas. The results in column (4) are based on the share of the population covered by an Low Emission Zone, accounting for heterogeneous population density in a catchment area using high resolution satellite population grids.[33]. The results are very similar to our main specification. This is also true if we apply an alternative approach by defining catchment areas by a ten minutes driving time radius around the hospital, shown in column (5).[34] Using binary indicators for being covered by an active Low Emission Zone in column (6) and (7) tend to increase point estimates. Given that this are rather broad definitions of treatment it not surprising that standard error increase.

Aggregating on municipality and county level. In order to reveal the importance of using high resolution spatial data we aggregate hospital diagnoses at the municipality and county level by year and regress the aggregate number of diagnoses on the share of a county or municipality that is covered by an active Low Emission Zone or on a binary indicator for an active Low Emission Zone. We use similar controls as in the main health regression but use the corresponding controls at the municipality or county level. The results are shown in Table 2.A.4.1 in Appendix 2.A.4. Columns (1) and (2) show the results on county level for the binary indicator and the share covered by a Low Emission Zone. Columns (3) and (4) show the corresponding results at the municipality level. While most of the coefficients are negative only two of them are marginally statistically significant. Furthermore, effect sizes are smaller in most of the cases compared to our main specification with high spatial resolution. This is additional evidence that the health benefits of Low Emission Zones are concentrated very locally for the the population living inside their boundaries.

Accounting for effects on traffic volume. Traffic is a potential additional channel when analyzing the impact of Low Emission Zones on health. If the implementation of Low Emission Zones reduces traffic volumes in addition to the vehicle fleet's emission standards there may be other impacts on public health in the long-term, for example on diseases of the circulatory system due to increased physical activity. Furthermore, less traffic could change the stress level in a city by lowering noise exposure or congestion. Based on a binary treatment indicator Table 2.A.5.1 in Appendix 2.A.5 shows the effect on traffic volume in and around a Low Emission Zone for all vehicles (columns (1)-(3)) and only passenger cars below 3.5 tonnes (columns (4)-(6)). In general, we control for the same characteristics as in

[33]We use the GEOSTAT 2006 and 2011 which are datasets of 1km $\times$ 1km population grids approximate by the building structure in each grid to calculate the population density in Low Emission Zones and in catchment areas. Our treatment is than the share of the population in a catchment area that is covered by a Low Emission Zone

[34]We chose ten minutes because three quarters of the German population reached the next general hospital within ten driving minutes in 2005 (Klauber, 2006).

our main specification. However, we now control for labor market region time trends instead of municipality time trends to account for changes in commuting behavior between municipalities. Most of the coefficients in Table 2.A.5.1 in Appendix 2.A.5 are negative but very small and none is statistically significant. These findings are in line with Wolff (2014) who shows that improvements in air quality are driven by increases in the share of low emitting vehicles in cities with Low Emission Zones.

Other diagnoses related to air pollution and traffic. Table 2.A.5.2 in Appendix 2.A.5 provides further evidence that health effects are driven by improvements in air quality through reductions in respiratory and cardiovascular diseases. Again, we use the same specification as in our main regression in Table 2.5 and study the effects on outcomes that may as well be affected by Low Emission Zones. For example, dementia and diabetes are suspected of being caused by air pollution. While we find negative point estimates for dementia, the results are not statistically significant. However, one would expect that improvements in air quality reduce the incidence of dementia only in the long run. We find no affect for diabetes. Additionally, we focus on stress related diagnoses and diagnoses of injuries which would reveal typical diagnoses related to road traffic accidents. Again, the coefficients are not significant which is line with the result that traffic volumes are not affected and suggest that air quality is the main driving factor for health improvements in Low Emission Zones.

Table 2.7: The effect of Low Emission Zones on diagnoses (alternative specifications)

Catchment area based on	Driving time			Driving time (POP)	10 min radius	Binary(in/out)	Binary(Covered)
Hospital sample	General Hospitals	Special hospitals	General hospitals in CAP	General hospitals	General hospitals	General hospitals	General hospitals
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All diseases (A00-N99)	-0.014*	-0.011	-0.030*	-0.010	-0.009	-0.015	-0.012
	(0.008)	(0.016)	(0.018)	(0.008)	(0.008)	(0.010)	(0.026)
Diseases of the ciculatory system (I00-I99)	-0.028***	-0.017	-0.087**	-0.022*	-0.022*	-0.030*	-0.026
	(0.011)	(0.042)	(0.043)	(0.012)	(0.011)	(0.017)	(0.044)
Hypertenion (I10-I15)	-0.031*	0.024	-0.044	-0.037*	-0.029*	-0.051***	-0.047
	(0.018)	(0.032)	(0.031)	(0.019)	(0.015)	(0.019)	(0.060)
Ischemic heart diseases (I20-I25)	-0.053***	0.005	-0.078*	-0.047**	-0.028	-0.053*	0.013
	(0.020)	(0.029)	(0.045)	(0.021)	(0.017)	(0.028)	(0.076)
Cerebrovascular disease (I60-I69)	-0.041**	-0.010	-0.075***	-0.034*	-0.040**	-0.033	-0.027
	(0.020)	(0.033)	(0.025)	(0.019)	(0.020)	(0.021)	(0.051)
Diseases of the respiratory system (J00-J99)	-0.021*	-0.022	-0.038	-0.014	-0.019	-0.022	-0.011
	(0.012)	(0.055)	(0.043)	(0.011)	(0.013)	(0.015)	(0.039)
Acute lower respiratory diseases (J20-J22)	-0.034*	-0.023	-0.057	-0.034*	-0.017	-0.048**	-0.117*
	(0.018)	(0.047)	(0.048)	(0.017)	(0.019)	(0.023)	(0.066)
Chronic lower respiratory diseases (J40-J47)	-0.045***	-0.010	-0.050	-0.044***	-0.039***	-0.063***	-0.088*
	(0.015)	(0.030)	(0.039)	(0.016)	(0.014)	(0.021)	(0.050)
Low birth weight (P07)	-0.018	0.015	0.014	-0.014	-0.024	-0.009	0.000
	(0.022)	(0.017)	(0.028)	(0.025)	(0.023)	(0.019)	(0.056)
N	8828	6874	5590	8828	8828	8828	8828
Controls:							
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linear municipality time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes:This table displays the results for different treatment definitions and samples. Each coefficient is the result of a separate regression of standardized diagnose listed on the left on a indicator variable for an active LEZ, while controlling for hospital and year fixed effects as well as hospital characteristics as well as hospital characteristics (non-profit, public, private, baserate, number of beds, number of beds^2), hospital size (small, medium, large) $\times$ years, municipality characteristics (mean temperature, precipitation and wind speed, population, work force, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)), linear time trends (Municipality $\times$ Years). Standard errors are clustered at county level and displayed in parentheses. Significance levels: $***: p < 0.01, **: p < 0.05, *: p < 0.1.$.

2.4 Conclusion

In this chapter, we show that Low Emission Zones are an effective policy instrument to reduce levels of air pollution in a targeted area, thereby having positive impacts on population health. Exploiting variation in the roll out of Low Emission Zones in Germany, we find that hospitals which catchment areas are covered by a Low Emission Zone diagnose significantly less air pollution related diseases. We find the effect to be stronger before 2012, which is consistent with a general improvement in the vehicle fleet's emission standards. Using precise spatial data on the extension of Low Emission Zones in Germany, our results confirm former results showing that the introduction of Low Emission Zones improved air quality significantly by reducing NO2 and PM10 concentrations. While effect sizes for average pollution levels are equal, our effect sizes for violations of air quality standards are larger compared to previous results (Wolff, 2014; Malina and Scheffler, 2015; Gehrsitz, 2017). This can be explained by our finding of a strong spatial delineation not captured by studies which use between and not within city variation as we do.

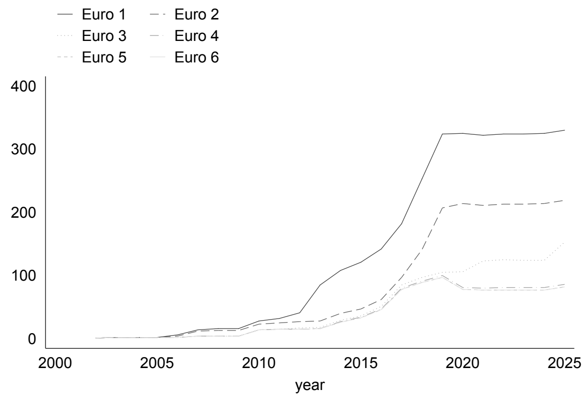
We show that the introduction of Low Emission Zones in Germany actually improved population health, in particular by reducing the incidents of chronic diseases of the circulatory and the respiratory system. Our results further suggest that these effects may be driven by reductions in non-emergency diagnoses of chronic diseases rather than emergency cases. We do not find reductions for low birth weight which is in line with Gehrsitz (2017). We further show that traffic volumes and diseases related to traffic (stress, injuries) were not affected by Low Emission Zones.

These findings have strong implications for policy makers. First, in 2015, overall costs for health care in Germany were around 340 billion euros, of which 46 billion euros for diseases of the circulatory system, making it the most expensive type of disease caused by 2.9 million cases (Statistisches Bundesamt, 2017b). Hence, reductions in the incidence of diseases of the circulatory system may directly reduce society's health costs. Besides, improving population health has sizable indirect costs on human capital and growth (Graff Zivin and Neidell, 2013). Second, the results of this chapter are informative for policy debates about further regulation of emissions from traffic. While the introduction of Low Emission Zones has reduced air pollution there are still numerous violations of EU air quality standards in German cities. As a consequence, as of 2019, vehicles with emission standards below Euro 5 or even Euro 6 (especially Diesel-fueled vehicles) are not allowed to enter designated areas in a number of large German cities (among others Stuttgart, Hamburg, Berlin and Cologne). These Diesel driving bans are currently controversially debated. Opponents question the potential health effects of these policy measures. While our findings show that restricting entry by high-emission vehicles improves population health through better air quality in inner-cities our findings are based on the regulation of emission standards Euro

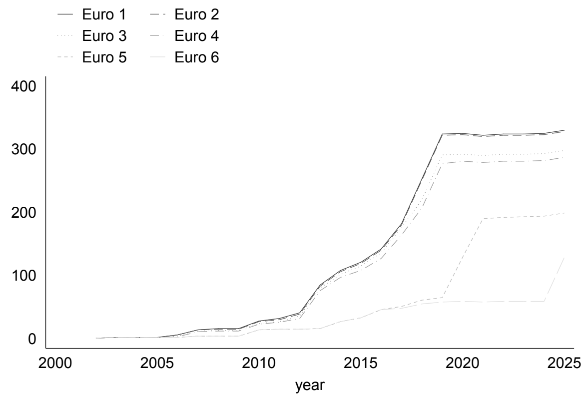
1–3. Whether further regulation of Euro 5–6 yields further health improvements should be addressed by future research.

APPENDIX

2.A.1 Low Emission Zones and air pollution in Germany



(a) Petrol



(b) Diesel

Figure 2.A.1.1 : Low Emission Zones in Europe

Notes: This figure shows the past and future development of Low Emission Zones across the European Union. Euro 2, 3, 4, 5 and 5 are subsets of Euro 1. Panel 2.A.1.1 a shows restrictions on Petrol vehicles and and Panel 2.A.1.1 a on Diesels. Source: urbanaccessregulations.eu.

Table 2.A.1.1: European Union air quality standards (PM10 and NO2)

(Pollutant)	(Thresholds)	(Deadline)
PM10	Yearly average limit 40$\mu g/m^3$ Daily average limit 50$\mu g/m^3$ Allowed number of transgression: 35	1 January 2005
NO2	Yearly average limit 40$\mu g/m^3$ Hourly average limit 200$\mu g/m^3$ Allowed number of transgression: 18	1 January 2010

Notes: This table displays air quality standards based on the Council Directive 1999/30/EC of 22 April 1999 relating to limit values for sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air. It was repealed by the Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008.

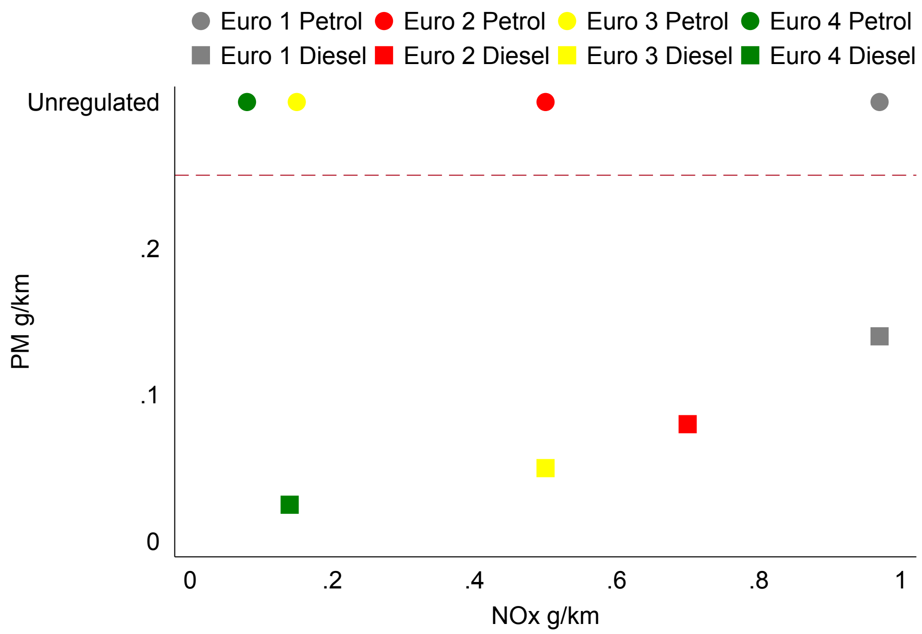


Figure 2.A.1.2 : Vehicle emission standards

Notes: This graph displays European emission standards of acceptable limits for exhaust emissions of new vehicles sold in the European Union and European Economic Area member states. They are defined in a series of European Union directives over time with increasingly stringent standards. Source: Tiwary and Williams (2018)

Table 2.A.1.2: Low Emission Zones in Germany as of 2018

(Low Emission Zone)	(Federal State)	(Sticker)	(Active since)	(Size in Km^2)	(Perimeter in Km)
Balingen	BW	Green	01.04.2017	90 Km^2	50 Km
Freiburg	BW	Green	01.01.2010	25 Km^2	58 Km
Heidelberg	BW	Green	01.01.2010	10 Km^2	34 Km
Heidenheim	BW	Green	01.01.2012	17 Km^2	28 Km
Heilbronn	BW	Green	01.01.2009	38 Km^2	28 Km
Herrenberg	BW	Green	01.01.2009	4 Km^2	9 Km
Ilsfeld	BW	Green	01.03.2008	2 Km^2	5 Km
Karlsruhe	BW	Green	01.01.2009	11 Km^2	16 Km
Leonberg / Hemmingen	BW	Green	02.12.2013	131 Km^2	60 Km
Ludwigsburg	BW	Green	01.01.2013	139 Km^2	58 Km
Möhlacker	BW	Green	01.01.2009	1 Km^2	7 Km
Mannheim	BW	Green	01.03.2008	7 Km^2	16 Km
Pfinztal	BW	Green	01.01.2010	31 Km^2	30 Km
Pforzheim	BW	Green	01.01.2009	2 Km^2	9 Km
Reutlingen	BW	Green	01.01.2008	109 Km^2	91 Km
Schramberg	BW	Green	01.07.2013	4 Km^2	16 Km
Schwäbisch Gmuend	BW	Green	01.03.2008	6 Km^2	17 Km
Stuttgart	BW	Green	01.03.2008	204 Km^2	109 Km
Tübingen	BW	Green	01.03.2008	108 Km^2	73 Km
Ulm	BW	Green	01.01.2009	28 Km^2	26 Km
Urbach	BW	Green	01.01.2012	2 Km^2	8 Km
Wendlingen	BW	Green	02.04.2013	4 Km^2	9 Km
Augsburg	BY	Green	01.07.2009	6 Km^2	12 Km
Müchen	BY	Green	01.10.2008	43 Km^2	28 Km
Neu-Ulm	BY	Yellow	01.11.2009	2 Km^2	21 Km
Regensburg	BY	Green	15.01.2018	1 Km^2	7 Km
Berlin	B	Green	01.01.2008	87 Km^2	38 Km
Bremen	HB	Green	01.01.2009	7 Km^2	13 Km
Darmstadt	HE	Green	01.11.2015	106 Km^2	90 Km
Frankfurt a.M.	HE	Green	01.10.2008	98 Km^2	60 Km
Limburg an der Lahn	HE	Green	31.01.2018	6 Km^2	15 Km
Marburg	HE	Green	01.04.2016	15 Km^2	34 Km
Offenbach	HE	Green	01.01.2015	39 Km^2	35 Km
Wiesbaden	HE	Green	01.02.2013	63 Km^2	78 Km
Hannover	NI	Green	01.01.2008	43 Km^2	30 Km
Osnabrück	NI	Green	04.01.2010	17 Km^2	33 Km
Aachen	NW	Green	01.02.2016	24 Km^2	28 Km
Bonn	NW	Green	01.01.2010	9 Km^2	18 Km
Düsseldorf	NW	Green	15.02.2009	43 Km^2	16 Km
Dinslaken	NW	Green	01.07.2011	4 Km^2	9 Km
Eschweiler	NW	Green	01.06.2016	2 Km^2	7 Km
Hagen	NW	Green	01.01.2012	9 Km^2	19 Km
Köln	NW	Green	01.01.2008	94 Km^2	88 Km
Krefeld	NW	Green	01.01.2011	10 Km^2	16 Km
Langenfeld	NW	Green	01.01.2013	1 Km^2	6 Km
Mönchengladbach	NW	Green	01.01.2013	21 Km^2	26 Km
Münster	NW	Green	01.01.2010	1 Km^2	6 Km
Neuss	NW	Green	15.02.2010	2 Km^2	6 Km
Overath	NW	Green	01.10.2017	1 Km^2	3 Km
Remscheid	NW	Green	01.01.2013	1 Km^2	7 Km
Ruhrgebiet	NW	Green	01.01.2012	868 Km^2	276 Km
Siegen	NW	Green	01.01.2015	3 Km^2	11 Km
Wuppertal	NW	Green	15.02.2009	25 Km^2	48 Km
Mainz	RP	Green	01.02.2013	34 Km^2	35 Km
Leipzig	SN	Green	01.03.2011	182 Km^2	111 Km
Halle (Saale)	SA	Green	01.09.2011	7 Km^2	12 Km
Magdeburg	SA	Green	01.09.2011	7 Km^2	21 Km
Erfurt	TH	Green	01.10.2012	16 Km^2	19 Km
Mean				49.96 Km^2	35.62 Km
Median				12.50 Km^2	21.31 Km
SD				119.39 Km^2	42.28 Km

Notes: This table shows detailed information of all active German Low Emission Zones in 2018. Source: OpenStreetMap.org., Federal Environment Office

(a) Official map (b) Own presentation

Figure 2.A.1.3 : Low Emission Zone of the Ruhr area

Notes: Panel (a) displays the LEZ of the Ruhr area based on official documents while Panel (b) shows the same LEZ based on polygons available at OpenStreetMap.org.

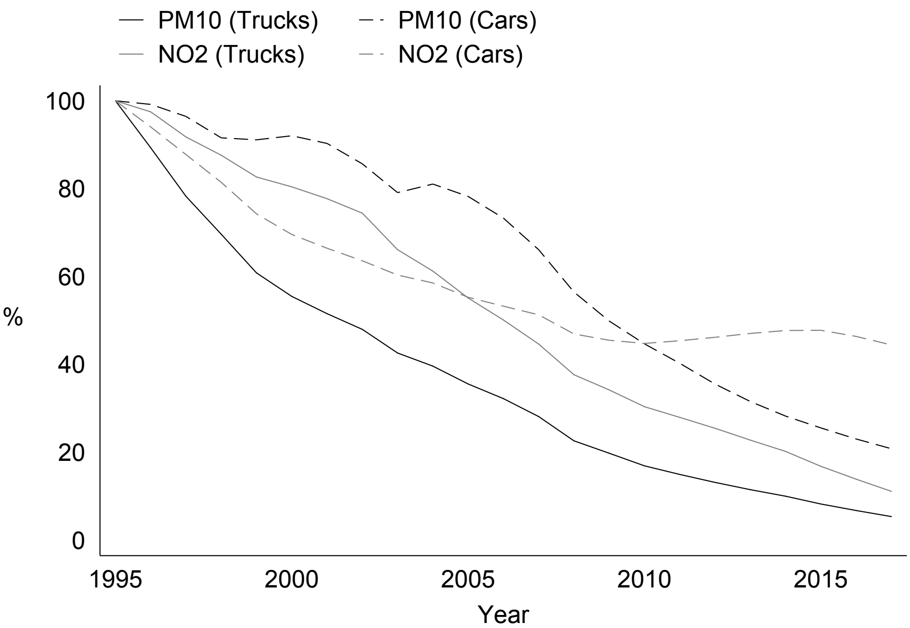
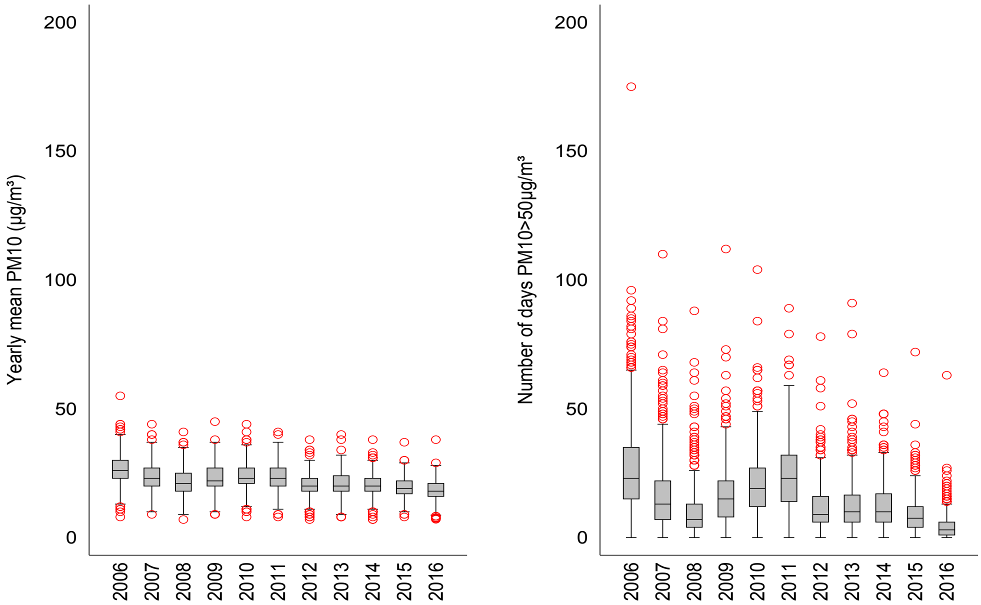
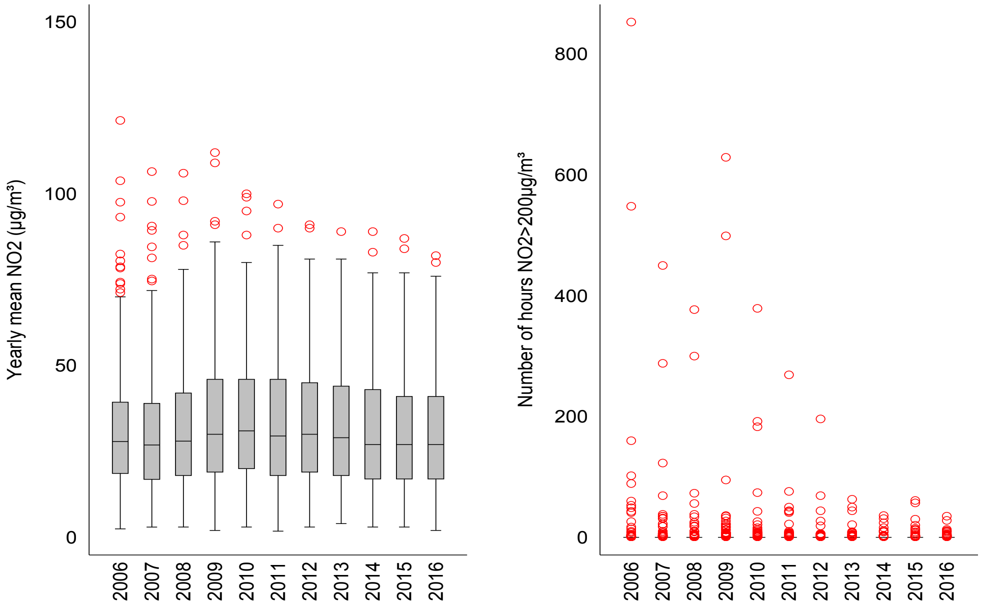


Figure 2.A.1.4 : Average emissions of vehicles in Germany

Notes: This Figure displays the development of of average vehicle emissions over time. Source: German Environment Agency.



(a) PM10

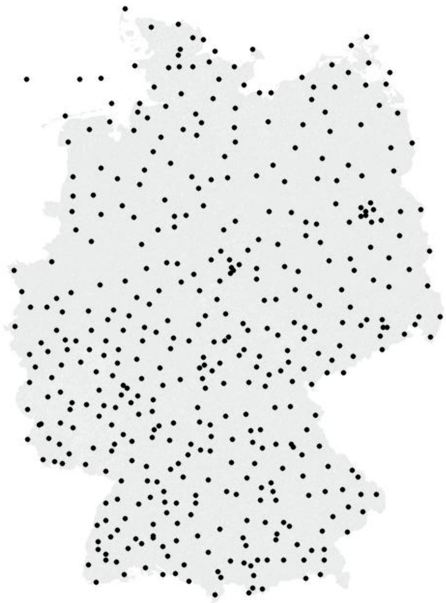


(b) NO2

Figure 2.A.1.5 : Variation of pollutants over time

Notes: These Figures display the yearly variation of pollutants over time. Source: German Environment Agency.

(a) Pollution monitors



(b) Weather monitors

Figure 2.A.1.6 : Location of pollution and weather monitors

Notes: These Figures display the measurement locations of pollution and weather monitors across Germany. Source: German Environment Agency, German Meteorological Service.

Table 2.A.1.3: The effect of Clean Air Plans on air pollution

	PM10		NO2	
	(1)	(2)	(3)	(4)
A. Pollution levels	Yearly mean PM10 ($\mu g/m^3$)		Yearly mean NO2 ($\mu g/m^3$)	
Clean Air Plan	-0.598***	-0.315	-0.598***	-0.162
	(0.206)	(0.212)	(0.278)	(0.296)
Clean Air Plan $\times$ In LEZ		-2.766***		-2.662***
		(0.546)		(0.546)
Adj. R^2	0.93	0.93	0.74	0.74
N	4290	4290	5237	5237
B. Limit exceedances	Yearly days PM10 > 50 ($\mu g/m^3$)		Yearly hours NO2 > 200 ($\mu g/m^3$)	
Clean Air Plan	-3.502***	-2.193**	4.376	5.501
	(0.824)	(0.859)	(3.175)	(3.813)
Clean Air Plan $\times$ In LEZ		-8.088***		-3.268
		(1.948)		(3.868)
Adj. R^2	0.81	0.82	0.50	0.50
N	4290	4290	4357	4357
C. Violations	Yearly mean PM10 > 40 ($\mu g/m^3$)		Yearly mean NO2 > 40 ($\mu g/m^3$)	
Clean Air Plan	0.010*	0.010	-0.008	0.002
	(0.006)	(0.007)	(0.020)	(0.019)
Clean Air Plan $\times$ In LEZ		-0.005		-0.027
		(0.005)		(0.049)
Adj. R^2	0.17	0.17	0.86	0.86
N	4290	4290	5237	5237
Controls:				
Station FE	Yes	Yes	Yes	Yes
State $\times$ Year FE	Yes	Yes	Yes	Yes
Weather characteristics	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes

Notes: Each column reports the result from a regression of the pollutant listed at the top on the treatment listed on the left, while controlling for monitor and year fixed effects as well as federal state time trends, weather characteristics (mean temperature, precipitation and wind speed) and municipality characteristics (population, workforce, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)). Standard errors are clustered at county level are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$..

Table 2.A.1.4: The effect of Low Emission Zones on air pollution by emission standard

	PM10			NO2		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Pollution levels	Yearly mean PM10 ($\mu g/m^3$)			Yearly mean NO2 ($\mu g/m^3$)		
In LEZ	-1.273***	-0.728***	-0.837***	-1.581***	0.577	0.056
	(0.204)	(0.210)	(0.207)	(0.460)	(0.522)	(0.466)
In LEZ × Euro 2		-0.831***			-3.116***	
		(0.241)			(0.724)	
In LEZ × Euro 3			-0.810***			-2.874***
			(0.223)			(0.654)
Adj. R^2	0.93	0.93	0.93	0.74	0.74	0.74
N	4290	4290	4290	5237	5237	5237
B. Limit exceedances	Yearly days PM10 > 50 ($\mu g/m^3$)			Yearly hours NO2 > 200 ($\mu g/m^3$)		
In LEZ	-6.580***	-3.934***	-4.031***	-5.572	1.582	-1.443
	(0.970)	(1.165)	(1.068)	(3.878)	(1.366)	(4.125)
In LEZ × Euro 2		-4.032***			-10.098*	
		(1.289)			(5.898)	
In LEZ × Euro 3			-4.735***			-7.147
			(1.114)			(5.357)
Adj. R^2	0.82	0.82	0.82	0.50	0.50	0.50
N	4290	4290	4290	4357	4357	4357
C. Violations	Yearly mean PM10 > 40 ($\mu g/m^3$)			Yearly mean NO2 > 40 ($\mu g/m^3$)		
In LEZ	-0.000	0.009	0.006	-0.043**	0.001	-0.022
	(0.006)	(0.009)	(0.009)	(0.022)	(0.031)	(0.026)
In LEZ × Euro 2		-0.015			-0.064	
		(0.009)			(0.030)	
In LEZ × Euro 3			-0.012			-0.037
			(0.008)			(0.027)
Adj. R^2	0.17	0.18	0.18	0.86	0.86	0.86
N	4290	4290	4290	5237	5237	5237
Controls:						
Station FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each coefficient is the result of a separate regression of the pollutant listed at the top on the treatment listed on the left while controlling for monitor and year fixed effects as well as federal state time trends, weather characteristics (mean temperature, precipitation and wind speed) and municipality characteristics (population, workforce, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)). Standard errors are clustered at county level are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$. .

Table 2.A.1.5: The effect of Low Emission Zones on air pollution in surrounding areas

	PM10			NO2		
	(1)	(2)	(3)	(4)	(5)	(6))
A. Pollution levels	Yearly mean PM10 ($\mu g/m^3$)			Yearly mean NO2 ($\mu g/m^3$)		
In LEZ	-1.273***	-1.229***	-1.181***	-1.581***	-1.527***	-1.512***
	(0.204)	(0.202)	(0.197)	(0.460)	(0.457)	(0.458)
10 km around LEZ		0.236	0.292		0.386	0.408
		(0.229)	(0.232)		(0.490)	(0.511)
10-20 km around LEZ			0.805***			0.297
			(0.281)			(0.690)
Adj. R^2	0.93	0.93	0.93	0.74	0.74	0.74
N	4290	4290	4290	5237	5237	5237
B. Limit exceedances	Yearly days PM10 > 50 ($\mu g/m^3$)			Yearly hours NO2 > 200 ($\mu g/m^3$)		
In LEZ	-6.580***	-6.359***	-6.209***	-5.572	-4.832	-4.577
	(0.970)	(0.934)	(0.922)	(3.878)	(3.416)	(3.291)
10 km around LEZ		1.170	1.345		4.333	4.669
		(0.866)	(0.880)		(3.265)	(3.441)
10-20 km around LEZ			2.538**			4.305
			(1.196)			(2.799)
Adj. R^2	0.82	0.82	0.82	0.50	0.50	0.50
N	4290	4290	4290	4357	4357	4357
C. Violations	Yearly mean PM10 > 40 ($\mu g/m^3$)			Yearly mean NO2 > 40 ($\mu g/m^3$)		
In LEZ	-0.000	-0.000	0.000	-0.043**	-0.045**	-0.045**
	(0.006)	(0.006)	(0.005)	(0.022)	(0.022)	(0.022)
10 km around LEZ		0.000	0.001		-0.012	-0.013
		(0.006)	(0.010)		(0.022)	(0.024)
10-20 km around LEZ			0.006*			-0.011
			(0.004)			(0.028)
Adj. R^2	0.17	0.17	0.17	0.86	0.86	0.86
N	4290	4290	4290	5237	5237	5237
Controls:						
Station FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column reports the result from a regression of the pollutant listed at the top on the treatment listed on the left, while controlling for monitor and year fixed effects well as federal state time trends, weather characteristics (mean temperature, precipitation and wind speed) and municipality characteristics (population, workforce, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)). Column (2) (4) and (6) report the results from a regression of the pollutant on a full interaction between the active LEZ and mutually exclusive group indicators. Standard errors are clustered at municipality level are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

2.A.2 Hospital Data

Hospital quality reports

Hospital quality reports are composed by hospitals and transferred to the Federal Joint Committee (Gemeinsamer Bundesausschuss) which collects and provides reports for the period 2006-2016. The Federal Joint Committee is a supreme decision-making body of the joint self-administration of physicians, dentists, psychotherapists, hospitals, and health care funds in Germany. The Federal Joint Committee, private health insurances, the German Medical Council (Bundesärztekammer) and the representative organizations of nursing professions are responsible for the content and extent of reports (6 § 137 SGB V) (Selbmann, 2004). Starting in 2004, hospitals were obliged to publish quality reports. However, only from 2006 onward reports were standardized and collected by the Federal Joint Committee. Reports are subdivided into hospital locations and hospital departments. The obligation to report refers to hospitals, hospital location, medicine departments that at least operated until 30. September of the reporting year. If closed before, no report is necessary. All provided information refer to the reporting year. Closing date is the 31. December of each year.

It is obligated to provide one report for one hospital location. A hospital location is legally defined in § 2a sec. 1 KHG (Krankenhausfinanzierungsgesetz), emphasizing the spatial and organizational independence. Building complexes with a linear distance not bigger than 2,000 meters can be defined as one location. Thus, if hospitals report several locations within a radius of 2,000 meter around the main location, which we define as the location with the highest initial number of inpatient cases, we merge these hospital locations. This happens 380 times. Otherwise, we would define competing hospital catchment areas for one hospital.

In order to calculate catchment areas, we need the geographic coordinates for each hospital location. We use the full addresses available in the quality reports and convert them using Nokias geocodingHere! API. This involves the input of the hospital address and a street network file provided by navteq for which an iterative comparison of the hospital address to the street network generates geographic coordinates. The calculation is based on interpolation along a street segment for which the geographic coordinates of the beginning and end points are known.

Quality reports are based on inpatient cases which are covered by the following funding schemes: Krankenhausentgeltgesetzes (KHEntgG) and Bundespflegesatzverordnung (BPflV). The BPflV covers a relative narrow scope, mainly treatments in psychological departments. The KHEntgG regulates the G-DRG fixed sum payment system which covers all diseases not covered by the BPflV. In combination, both system cover all in-patient cases. Diagnoses we are using are using for our analysis are based on the (KHEntgG). Under the

KHEntgG scheme, one case equals one diagnose in the year of dismissal. Different than under the KHEntgG system, reallocation of patients between medical departments increase the number of inpatient cases under the BPflV scheme. Thus, the number of in-patients can differ from the number of main diagnoses. Readmission does not increase the number of inpatient cases under both funding schemes.

The number of main diagnoses that we use as our identifier for population health is based on the German coding references (ICD-10-GM). The ICD-10-GM is an adaptation of ICD-10-WHO, the World Health Organization's "International Statistical Classification of Diseases and Related Health Problems". It is translated into German by the German Institute of Medical Documentation and Information (DIMDI). Main diagnoses are provided at 4 digit level. The main diagnose is defined as the disease primarily responsible for in-patient hospitalization. Due to data protection, diagnoses with less than six patients per year equal five.

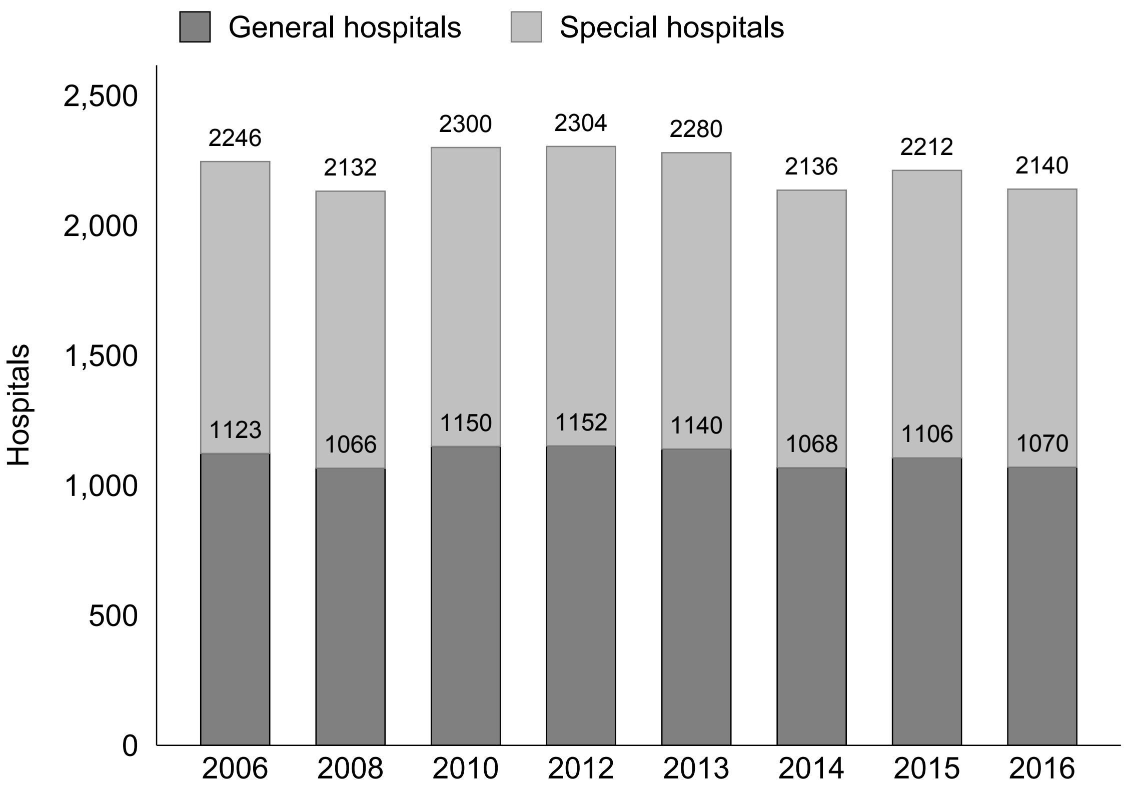


Figure 2.A.2.1 : Number of hospital locations

Notes: This figure shows the number of all German hospital locations separated by general and special hospitals

Catchment areas

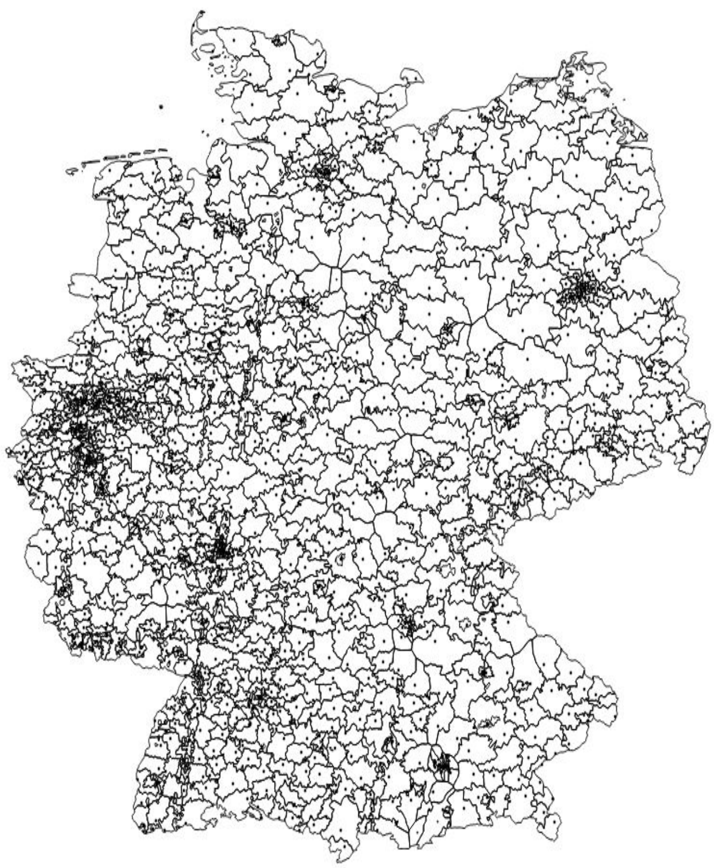


Figure 2.A.2.2 : Hospital locations and catchment areas (all Germany)

Notes: This figure displays all hospital locations and their catchment areas as of 2006 based on driving time.

Figure 2.A.2.3 : Hospital locations and catchment areas (Bonn)

Notes: This graph displays the Low Emission Zone in the city of Bonn (dark colored area: initial zone implemented in 2010, light area: extension as of 2012) as well as hospital locations (black dots) and their corresponding catchment areas (black lines) based on driving time.

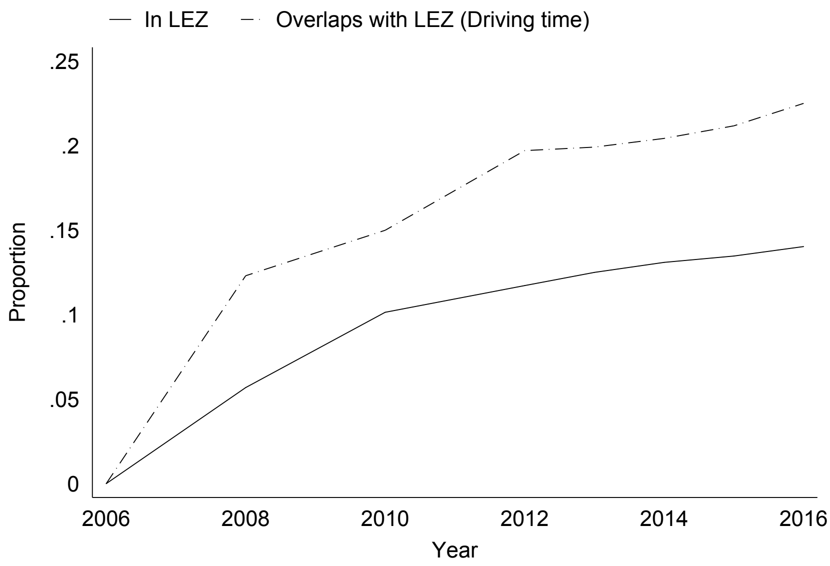
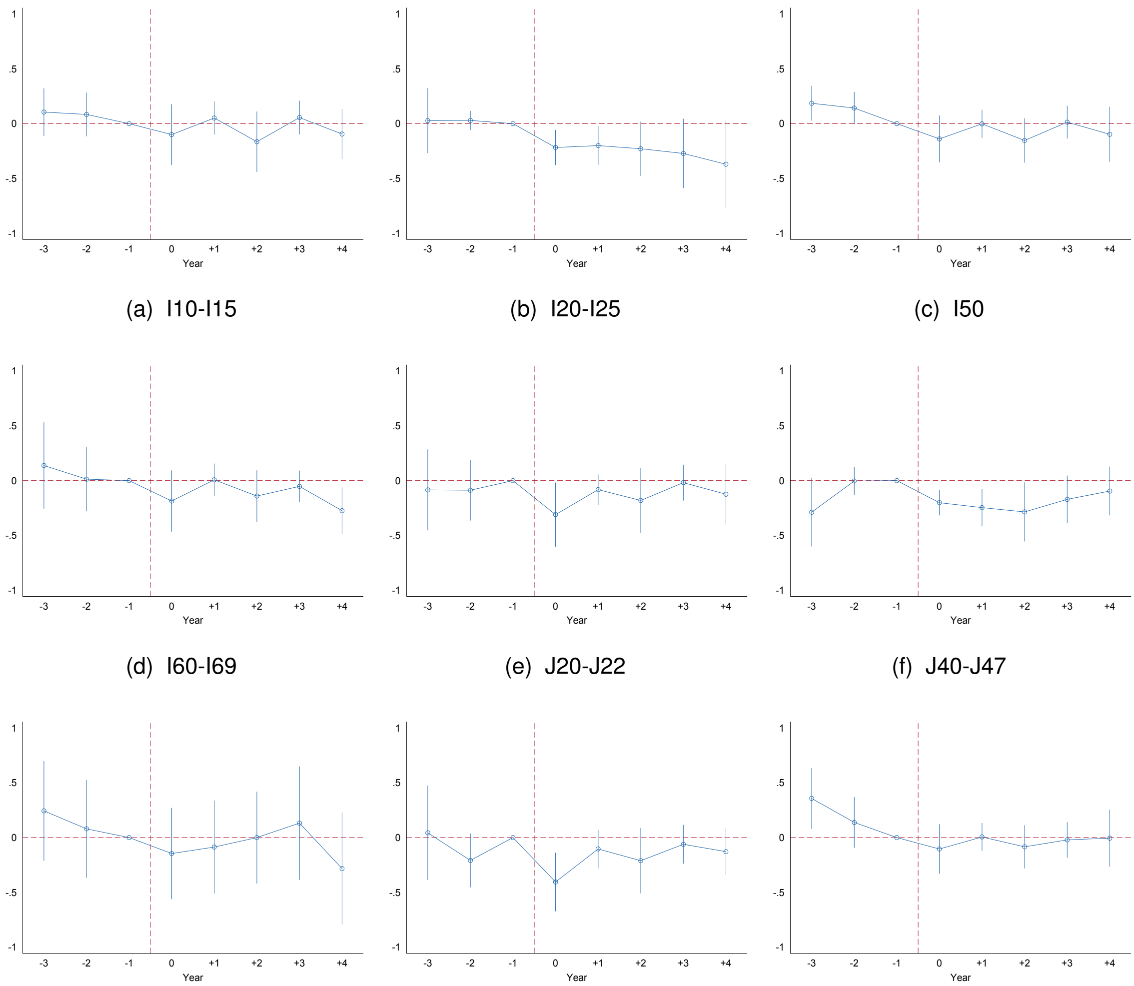


Figure 2.A.2.4 : Proportion of hospitals covered by Low Emission Zones

Notes: This graph displays time trends for the share of hospitals that are either located in an active LEZ or have catchment areas covered by an active LEZ.

2.A.3 Additional Results



(a) I10-I15 (b) I20-I25 (c) I50

(d) I60-I69 (e) J20-J22 (f) J40-J47

(g) P07 (h) F40-F48 (i) S00-S99

Figure 2.A.3.1 : Event study for all main diagnoses

Notes: These figures display event studies revealing the impact of $\beta\ shareLEZ_{it}$ on all main diagnoses). The reference period is k = -1. Each coefficient is the result of a separate interactions of dummy variables counting the years before and after the introduction of an LEZ and an indicator variable showing if the share of a hospital catchment area covered by an active LEZ, while controlling for hospital and year fixed effects as well as federal state time trends, hospital characteristics (non-profit, public, private, baserate, number of beds, number of beds^2), hospital size (small, medium, large) $\times$ years, weather characteristics (mean temperature, precipitation and wind speed) and municipality characteristics (population, workforce, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)) and linear municipality time trends. Standard errors are clustered at county level

2.A.4 County-level analysis

Table 2.A.4.1: The effect of Low Emission Zones on diagnoses on the county level

	County-level		Municipality-level	
	(Binary)	(Share)	(Binary)	(Share)
	(1)	(2)	(3)	(4)
All diseases (A00-N99)	-0.014	0.026	0.017	0.000
	(0.034)	(0.078)	(0.033)	(0.078)
Diseases of the ciculatory system (I00-I99)	-0.041	-0.123	0.046	-0.100
	(0.062)	(0.135)	(0.058)	(0.129)
Hypertension (I10-I15)	-0.066	-0.232	0.063	-0.084
	(0.062)	(0.147)	(0.101)	(0.187)
Ischemic heart diseases (I20-I25)	-0.026	-0.019	0.118	0.109
	(0.093)	(0.179)	(0.107)	(0.219)
Cerebrovascular disease (I60-I69)	-0.062	-0.321	-0.020	-0.375*
	(0.063)	(0.240)	(0.072)	(0.219)
Diseases of the respiratory system (J00-J99)	0.035	0.008	0.048	0.024
	(0.051)	(0.150)	(0.057)	(0.148)
Acute lower respiratory diseases (J20-J22)	-0.030	0.014	-0.053	0.036
	(0.067)	(0.184)	(0.091)	(0.218)
Chronic lower respiratory diseases (J40-J47)	-0.010	-0.241*	0.028	-0.185
	(0.055)	(0.140)	(0.068)	(0.158)
Low birth weight (P07) [t+1]	0.071	-0.128	0.127	-0.021
	(0.111)	(0.271)	(0.105)	(0.247)
N	3024	3024	6292	6292

Notes: This table displays the results for hospital diagnoses, at the municipality and county level. Each coefficient is the result of a separate regression of diagnose listed on the left on a indicator variable for an active LEZ (share of municipality or county covered by LEZ, or a binary indicator being one if a municipality or county has an active LEZ and 0 otherwise), while controlling for municipality or county and year fixed effects as well as federal state time trends, municipality or county characteristics (mean temperature, precipitation and wind speed, population, work force, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)). Standard errors are clustered at unit level of observation and displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

2.A.5 Additional channels

Table 2.A.5.1: The effect of Low Emission Zones on traffic

	All vehicles			<3.5t		
	(1)	(2)	(3)	(4)	(5)	(6)
In and 10 km around LEZ	-0.002	-0.003	-0.003	-0.003	-0.004	-0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
10-20 km km around LEZ		-0.010	-0.009		-0.011	-0.010
		(0.006)	(0.006)		(0.007)	(0.007)
20-30 km around LEZ			0.006			0.007
			(0.013)			(0.013)
Adj. R^2	0.23	0.23	0.23	0.21	0.21	0.21
N	12052	12052	12052	12052	12052	12052
Controls:						
Monitor FE	Yes	Yes	Yes	Yes	Yes	Yes
LMR $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Weather characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Notes:Notes: Each column reports the result from a regression of traffic volume on the treatment listed on the left, while controlling for monitor and year fixed effects well as labor market region (LMR) time trends, weather characteristics (mean temperature, precipitation and wind speed) and municipality characteristics (population, workforce, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)).. Standard errors are clustered at county level are displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$. .

Table 2.A.5.2: The effect of Low Emission Zones on further diagnoses in general hospitals

	(1)	(2)	(3)	(4)	(5)
Dementia (F00-F03)	-0.049	-0.085	-0.084	-0.129	-0.083
	(0.121)	(0.121)	(0.121)	(0.128)	(0.129)
Diabetes (E10-E14)	0.049	0.045	0.047	0.050	0.007
	(0.146)	(0.136)	(0.136)	(0.145)	(0.121)
Stress (F40-F48)	0.014	-0.007	-0.010	-0.007	-0.111
	(0.105)	(0.099)	(0.098)	(0.094)	(0.099)
Injuries (S00-S99)	0.016	-0.002	-0.001	-0.033	-0.145
	(0.067)	(0.066)	(0.066)	(0.074)	(0.104)
N	8828	8828	8828	8828	8828
Controls:					
Hospital FE	Yes	Yes	Yes	Yes	Yes
State × Year FE	Yes	Yes	Yes	Yes	Yes
Hospital characteristics	No	Yes	Yes	Yes	Yes
Weather characteristics	No	No	Yes	Yes	Yes
Municipality characteristics	No	No	No	Yes	Yes
Linear municipality time trends	No	No	No	Yes	Yes

Notes:This table displays the results for hospital diagnoses, for main hospitals. The catchment area is calculated by driving time. Each coefficient is the result of a separate regression of diagnose listed on the left on a indicator variable for an active LEZ (share of catchment area covered by LEZ), while controlling for hospital and year fixed effects as well as federal state time trends, hospital characteristics (non-profit, public, private, baserate, number of beds, number of beds^2), hospital size (small, medium, large) × years, municipality characteristics (mean temperature, precipitation and wind speed, population, work force, age structure (share men(min-30, 31-64, 65-max), women(min-30, 31-64, 65-max)), linear time trends (Municipality × Years). Standard errors are clustered at county level and displayed in parentheses. Significance levels: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

CHAPTER 3

Reallocation of Hospital Resources

Changes in medical expenses may force hospitals to reallocate their resources, which potentially come at the detriment of healthcare quality. Using data on the universe of German hospitals, I investigate resource reallocations between capital stock, human resources, services and the organizational structure in case of reform-induced treatment price shocks at the hospital level. To identify a causal effect, I develop a unique identification strategy where I exploit hospitals' exposure to snowfall. A particularity of the reform led to exogenous treatment price shocks at hospital level in response to weather-induced excess of patients. The results show that higher prices induce hospitals to hire more physicians and nurses and encourage fewer mergers and privatization and less closures while not affecting the capital stock. In addition, hospitals become less specialized and reduce their treatment volume. These effects persist long after the treatment price shocks have vanished. [1]

[1]The working paper version of this chapter was published as IZA Discussion Paper No. 13256

3.1 Introduction

In the past decades, healthcare spending in OECD countries has increased massively and reached 4,069 $ per capita in 2017, exceeding spending on education by 70 percent. Hospitals have become the main providers of healthcare services (OECD, 2018). At the same time, large-scale reallocation processes of hospital resources have been observed. The number of cases and treatments increases dramatically, while the number of input factors such as nursing staff remains relatively constant. Understanding the effect of changes in healthcare spending on the reallocation of hospital resources is important because hospital resources are crucial for the quality of healthcare (Needleman et al., 2002). However, the impact of healthcare spending on hospital resources has been debated in a so far inconclusive literature (Sloan and Hsieh, 2017).

If hospital healthcare spending changes, hospitals can adjust along three resource dimensions. First, they can adjust input factors such as nurses or physicians to save costs or increase productivity. Second, hospitals can adjust output factors such as the variety or volume of treatments, thereby realizing economies of scale or turnover increases. Third, hospitals could change their ownership structure by merging or privatizing to minimize transaction costs. In the worst case, they have to be closed (Guterman and Dobson, 1986).

In this chapter, I show how hospitals reallocated resources between 2006 and 2016 due to exogenous general treatment price shocks between 2005 and 2010. The analysis is based on panel data from the universe of German hospitals, which offers a unique abundance of hospital resource information. For identification, I use the introduction of the German Diagnosis Related Groups (G-DRG-System). Diagnosis Related Groups have been implemented worldwide (Busse and Quentin, 2011). The G-DRG-Reform led to general idiosyncratic treatment price shocks for hospitals, but did not alter relative treatment prices for different types of care or groups of patients. This is an advantage to reforms used in most of the literature, which are based on relative treatment price shocks (see, for example, Dafny (2005); Acemoglu and Finkelstein (2008); Clemens and Gottlieb (2014)). Furthermore, I rely on plausible exogenous variation in treatment prices, which is ensured by a novel identification strategy. I exploit days of snow in hospitals catchment areas, which led to exogenous treatment price shocks at hospital level in response to weather-induced excess of patients due to a particularity of the G-DRG-Reform. I focus on the reallocation of human resources, capital stock as well as changes in services offered and the organizational structure. This allows me to simultaneously analyze underlying hospital preferences for input and output factors under general treatment price shocks.

The implementation of the G-DRG-System in 2004 caused continuous idiosyncratic treatment price shocks at the hospital level between 2005 and 2010. Starting with hospital individual prices in 2004, prices converged to the federal state average until 2010. Some hospitals were exposed to price increases, while others experienced price reductions. The variation in price shocks was large. Due to strong price decreases, many hospitals encountered economic difficulties and an increasing risk of economic default, especially those in public ownership (Klauber et al., 2018). From 2010 onwards, treatment price shocks only varied at the federal state level, enabling a novel study of long-term effects of price shocks after price shocks vanish. The G-DRG-Reform enhanced treatment price decreases and mitigated treatment price increases between 2005 and 2010 for individual hospitals if its number of in-patient[2] cases in 2004 was higher and more severe than in 2003 compared with other hospitals in a federal state.

Treatment price rice shocks of the G-DRG-Reform are endogenous to the allocation of hospital resources, for several reasons. First, they reflect historic cost structures, which might correlate with resource allocations after the G-DRG-Reform. Second, hospitals were able to manipulate treatment price shocks in 2004 by shifting in-patient cases. In order to avoid that the results are driven by unobserved heterogeneity, I instrument treatment price shocks by exploiting idiosyncratic changes of weather conditions between 2003 and 2004 in hospital catchment areas. This approach is based on unique high-resolution satellite data of population-weighted exposure to the number of yearly days of snow and road network data, which I link to the hospital location. Unlike other datasets, the hospital panel data in this study has information about the exact hospital location. Days of snow have a strong positive impact on the number of cost-intensive hospital admissions, easily doubling the number of related admissions in winter months (Franklin et al., 1995). I focus on general hospitals, which usually treat snow related admissions. Germany is an interesting case for this novel identification strategy due to its changeable weather conditions, which makes it difficult for people to avoid accidents by adjusting their behavior. First-stage results reveal a strong correlation between changes in the number of days of snow and hospital individual treatment prices, driven by snow-related admissions.

The results of the chapter show that treatment price shocks significantly affect hospital resources. Effects of price shocks are linear, which is why the G-DRG-Reform leads to a polarization of healthcare. Treatment volumes are negatively and ranges of treatments are positively associated with price shocks. Hospitals significantly reallocate their resources towards smaller treatment volumes and an extended range of treatments if prices increase, which indicates supplier-induced demand. However, as the G-DRG-Reform reduced treatment prices for a large share of hospitals, supplier-induced demand counteracts saving affords and negatively affects public health. The correlation between treatment price shocks and the number of employed nurses and physicians is positive. Treatment price reductions decrease the nurse to patient ratio and the physician to patient ratio, possibly to the

[2]Treatments that require at least one overnight stay

detriment of patients' health (Aiken et al., 2002). The stock of capital is unaffected by treatment price shocks. These effects tend to be persistent, even if idiosyncratic treatment price shocks vanish after 2010. Structural differences in pre-treatment characteristics lead to heterogeneous effects. For instance, private and small hospitals are more strongly affected. Furthermore, treatment price shocks affect the organizational structure of a hospital by being negatively associated with the probability of mergers, privatization and closures. IV estimates show that OLS results are biased towards zero in almost all dimensions, indicating endogenous price shocks.

This chapter adds to several strands of the literature. By analyzing the effect of universal treatment price shocks on input factors, it adds to the literature of Finkelstein (2007), Acemoglu and Finkelstein (2008) and Clemens and Gottlieb (2014), finding responses to medical investment decisions for capital and human resources based on relative price shocks. Furthermore, by focusing on the range of treatments and the treatment volume, this chapter adds to the literature showing a clear link between relative price shocks and the range and volume of treatments (Rice, 1983; Yip, 1998; Dafny, 2005; Clemens and Gottlieb, 2014). By exploting the G-DRG-Reform as well, Salm and Wübker (2015, 2018) show that general treatment price shocks affected the number of treatments and input factors, while not affecting the quality of care. By examining changes in organizational structures, I complement the work of Krishnan (2001) and Gowrisankaran et al. (2015), who show a clear positive relationship between relative price increases and the probability of changes in the organizational structure of hospitals. Compared to these studies, I simultaneously focus on input and output factors and on the organizational structure while relying on plausibly exogenous variation in general treatment prices. Further, I analyze persistent patterns in the reallocation of hospital resources even when treatment price shocks vanish.

The remainder of this chapter is structured as follows. In section 3.2, I provide background information about the G-DGR-Reform and the datasets. Section 3.3 describes the empirical strategy, followed by the empirical analysis in section 3.4 Section 3.5 concludes.

3.2 Institutional Background and Data

3.2.1 The German Diagnosis Related Groups Reform

The implementation of the G-DRG-System caused massive changes in the way hospitals set their treatment prices. Like in many countries, German hospitals set their treatment prices based on the full costs of services in a fee-for-service payment system (FFS) with few restrictions. Whereas in DRG-Systems prices are fixed for given diagnoses, prices in FFS are not fixed and depend on the treatments a patient received. In the German FFS, at the end of each year, hospitals negotiated with relevant health insurance companies[3] about their next year's treatment prices, considering the present cost structure and the hospitals' supply mandate for the following year. In the following year, hospitals could deviate from the agreed service quantity. Additional costs were compensated if medically justified.

In contrast to other countries, the G-DRG-System was planned as a universal price system. The healthcare reform[4] in 2000 asked the relevant self-governing bodies at the federal level to implement an extensive DRG-System for hospitals until January 1, 2003. In 2003, hospitals optionally chose between the old FFS and the new DRG-System. From 2004 onwards, the new system was mandatory and it remains in place until today. However, the DRG-System in 2003 and 2004 was "budget-neutral" as the new and old systems coexisted, giving hospitals the possibility to become familiar with the new accounting procedures. The budget was still negotiated as before, although costs were accounted according to the new system. The G-DRG-System applies to almost all patients, regardless of whether or not they are members of the statutory health insurance system, private health insurance, or self-paying patients. It further applies to all hospitals including all clinical departments with the exception of institutions or facilities providing psychiatric, psychosomatic medicine or psychotherapy services (Fürstenberg et al., 2013). After the reform, more than 95 percent of the hospital budget was reimbursed according to the G-DRG-System (Klauber et al., 2011).

Treatment prices in the G-DRG-System are set as in other DRG-Systems. Since 2004, treatment prices under the G-DRG-System have been based on the following formula

$$price_{iht} = drg_{it} \times baserate_{ht}, \tag{3.1}$$

where drg_{it} (G-DRG) is the cost-weight factor for a diagnose i in year t, while $baserate_{ht}$ refers to the base rate for a hospital h in year t, which acts as a baseline. The Institute for the Hospital Remuneration System calculates levels for each G-DRG, representing the average estimated costs for given diagnoses at national level. G-DRGs are meant to cover the medical treatment, the provision of pharmaceuticals and therapeutic appliances, nursing care, food and accommodation. In contrast to other countries, G-DRGs do not cover capital costs. Officially, federal states should pay for investments (Herr et al., 2011).

The base rate in the G-DRG-System is set in a unique way. The aim of the reform was to equalize treatment prices within federal states. Usually, DRG-Systems in other countries account for structural hospital differences by hospital specific base rates, which equal yearly average treatment costs per case for each hospital (Schreyögg et al., 2006). Hospital specific base rates were only temporarily implemented in Germany and since 2010 only vary at

[3]With those health insurance companies that covered >= 5 percent of the cases in a hospital in the previous year.
[4]GKV-Gesundheitsreformgesetz

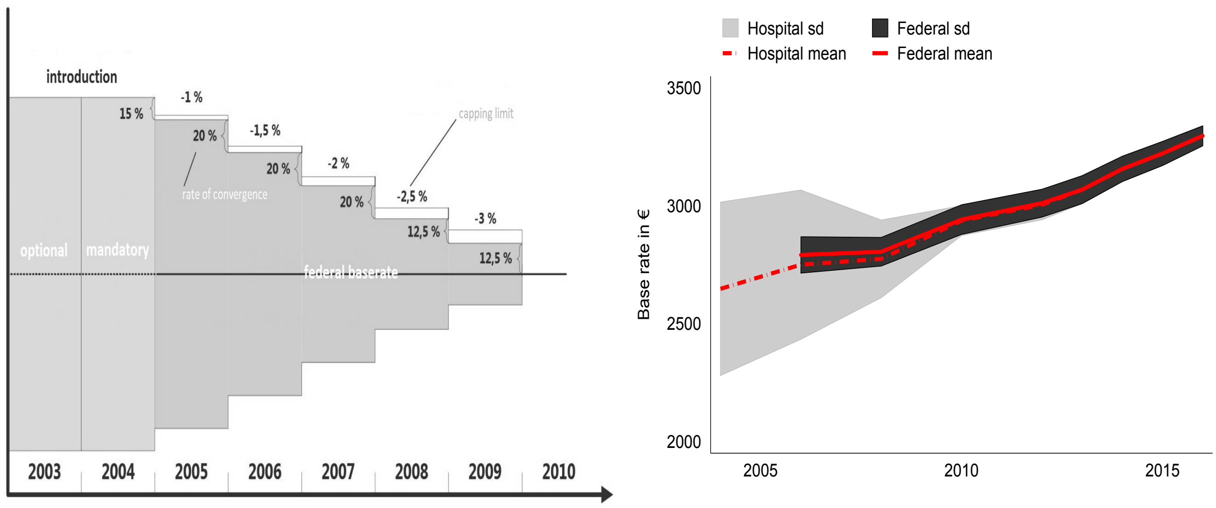
the federal state level. Hospital-specific base rates were calculated by dividing the hospital's 2004 budget – which was negotiated at the end of 2003 based on the old accounting system – by the 2004 case mix.[5] The case mix is part of the new accounting system, reflecting the yearly number of in-patient cases weighted by their severeness. In 2004, hospital base rates considerably varied, ranging from less than 1,000 € to more than 10,000 €, reflecting huge differences in hospital cost structures. Hospitals with different structures than the average in a federal state might not be able to replicate average treatment costs. In order to mitigate reform effects, a convergence phase was implemented from 2005 until 2010, in which hospital-specific base rates of 2004 converged to the average hospital base rate of the federal state (Figure 3.1a). The federal state-specific base rate was calculated for each year starting in 2004, based on the accumulated budgets of all hospitals in a federal state divided by their accumulated case mixes. Hospital base rates were then gradually reduced or increased to the federal base rate.[6] Thus, the G-DRG-Reform enhanced treatment price decreases and mitigated treatment price increases between 2005 and 2010 for an individual hospital if the number of inpatient cases in 2004 was higher an more severe than in 2003 compared with other hospitals in a federal state. Originally, the convergence phase should have ended in 2009 but was prolonged until 2010 as many hospitals faced financial difficulties. The development of the actual federal and hospital base rate is shown in Figure3.1b, indicating the effectiveness of the reform. From 2010 onwards, base rates only varied at the federal state level.

However, treatment price shocks in the G-DRG-System are potentially endogenous. The convergence phase was intended to allow hospitals to adjust structures over time. Thus, price shocks reflect historic resource allocations. For example, Figure 3.A.1.1 in Appendix 3.A.1 shows a strong correlation between the number of beds as well as the number of physicians with the hospital base rate of 2004. Lower base rates in 2004 are a signal of efficiency. However, more efficient hospitals might adjust hospital resources more strongly to treatment price shocks in the convergence period if they are more flexible than inefficient hospitals. This would bias the results towards zero. Furthermore, hospitals had an incentive in reducing the case mix of 2004 to increase their base rate, or in other words to improve their starting position of the convergence phase, regardless of whether they were above or below the federal base rate. In fact, Figure 3.A.1.2 in Appendix 3.A.1 shows a strong decline in in-patient cases in 2004 – which reduced the case mix – followed by an immediate recovery. Hospitals that adjust the number of in-patient cases more strongly in 2004,

[5]case mix = in-patient cases $\times$ case mix index. The case mix index is a hospital specific weighting factor for the average severeness of yearly admissions

[6]A capping limit between one and three percent of the hospitals' budget was implemented for those hospitals with decreasing base rates. Base rate harmonization could only lead to yearly reductions in hospital budgets within these capping limits. Such capping limits were not in place for hospitals whose base rates increased. This increased the overall hospital budged and the federal baserate (Fürstenberg et al., 2013).

might be those hospitals that adjust hospital resources stronger in the convergence phase. Estimates would be upward biased. Ex ante, it is unclear whether the results are upward biased by the adjustment of in-patient cases in 2004 or biased towards zero by correlations between previous resource allocation and resource reallocations in the convergence phase.



(a) Convergence phase

(b) Actual base rate

Figure 3.1: Base rate development over time

Notes: These graphs display (3.1a) the planned convergence of hospital base rates in the federal state and (3.1b) the actual convergence of hospital base rates cross Germany. Source: AOK

3.2.2 Mechanisms

This section highlights relevant hospital resources that are prone to external treatment price shocks. The selection of relevant resources is based on the assumption that hospitals try to maximize profits by adjusting input and output factors and the organizational structure to realize the efficient use of resources, higher intensity and improvements in productivity (Hodgkin and McGuire, 1994; Epstein and Mason, 2006; Sánchez-Martínez et al., 2006).

Input factors. The provision of hospital treatment involves utilizing a variety of different inputs during the production process, especially human resources and stock of capital. The most important input factor in a hospital is human resources, especially the number of physicians and nurses, with strong positive implications for hospitals productivity (Needleman et al., 2002; Unruh, 2003). However, labor is the most important cost factor, characterized by high variation between staff categories (Saltman and Figueras, 1998). This makes it very likely that human resources are positively associated with treatment price shocks. The same also applies to the stock of capital. Substantial capital investments hold increasing interest for medical suppliers (Levin et al., 2008) due to their strong positive impact on productivity. Capital-intensive investments like MRI scanners are likely to be amplified by treatment price shocks as they come with negligible marginal costs. (Clemens and Gottlieb, 2014).

Output factors. While input factors are major leverages, hospitals could try to increase their efficiency by portfolio adjustments towards higher treatment volumes and narrower treatment ranges to realize economies of scale (Farsi and Filippini, 2008). Medical providers can indirectly drive the quantity of their patients' healthcare due to the consumers' lack of information about treatment options (Arrow, 1978). The benefits of adhering to ethical and medical standards are weighted against the benefits of higher revenues, leading to the counter-intuitive prediction that medical providers will supply a reduced treatment volume and a narrower treatment range if treatment prices increase (Evans, 1974; McGuire, 2000). Especially, if health insurance diminishes or eliminates price sensitivity (Feldstein, 1973). As health insurance is statutory in Germany and fully covers hospital stays, it is reasonable that treatment price shocks do not directly affect demand but rather indirectly via supplier-induced demand.

Organizational structure. The hospital ownership is an important characteristic in the context of treatment price shocks, especially in Germany where many hospitals have been privatized in the last two decades. Public hospitals finance around 50 percent of their investments themselves. Officially federal states are responsible for such investments, although they are unable to comply with the duty due to financial distress. Moreover, many municipalities – which are typically the owners of public hospitals – are financially constrained (Augurzky et al., 2010). Hospitals can change the organizational structure as response to treatment price shocks to realize cost savings through mergers or privatization (Chalkley and Malcomson, 2000; Kristensen et al., 2010; Herr et al., 2011). If hospital produce at above average costs and do not realize cost savings or revenue increases, hospitals have to close (Klauber et al., 2018). This makes it likely that privatizations, mergers and closures are negatively associated with treatment price shocks.

3.2.3 Data and Descriptive Statistics

For the analysis of treatment price shocks on resource allocation, I link panel data of the universe of hospitals in Germany with high-resolution satellite data and routing data. In this section, I describe the construction of the dataset and present descriptive statistics. Most of the hospital dataset equals the dataset used in section 2. Thus, I limit the construction of the dataset to the parts that differ from section 2.

Hospital quality report data comprise hospital characteristics like the number of beds and ownership structure as well as detailed information about services offered, the stock of employees and special equipment between 2006 and 2016. A detailed explanation of the dataset can be found in section 2.2.3. I restrict the dataset to general hospitals with at least a unit for internal medicine and a surgery unit which are plausibly affected by snow induced demand shocks. Thus, I exclude special hospitals like hospices, wellness clinics, rehabilitation centers etc. resulting in around 1,100 hospitals per year (see Figure 2.A.2.1 in Appendix 2.A.2). I will use special hospitals for the robustness check. Furthermore, I can only use hospitals which already existed in 2004, resulting in 8,626 hospital-year observations. I supplement the quality reports with the German Hospital Directory[7] to have information about hospital locations and basic hospital characteristics before 2006, provided by the Federal Statistical Office. Having data about hospital locations before 2006 is essential for the IV, as it is based on hospital catchment areas in 2003 and 2004.

Base rate deviation. Information about hospital and federal base rates is provided by the Federal Association of the AOK, a representative of the interests of many local health insurance companies in Germany. The main regressor of interest $(br_{ft_{2004}} - br_{it_{2004}})$ is the deviation between the federal base rate in 2004 $(br_{ft_{2004}})$ and the hospital base rate in 2004 $(br_{it_{2004}})$, which determines treatment price shocks in the following years.

Hospital catchment areas are assigned based on hospitals' locations since the hospital quality report data does not provide information on the residence of patients. Catchment areas are based on driving time and calculated as in chapter two. A detailed explanation of the catchment area calculation can be found in section 2.2.3.

Snow data. Hospitals' exposure to days of snow is assigned by overlaying snow fall grids with the hospital catchment areas and the GEOSTAT population grid dataset. The snowfall data is part of the Global Snow Pack dataset and is provided by the German Aerospace Center. Since 2000, it has measured the daily existence of snow on the ground at a 500 $\times$ 500 meter resolution based on satellite imaging (Dietz et al., 2015). Figure 3.A.2.1 a in Appendix 3.A.2 shows a strong variation in days of snow over time, which is also true between consecutive years within catchment areas across Germany (Figure 3.A.2.1 b in Appendix 3.A.2). The GEOSTAT population grid dataset is provided by EUROSTAT and measures the population density at a 1,000 $\times$ 1,000 meter resolution based on satellite imaging in 2006 and 2011.

To have an intuition on how the instrument is constructed, Figure 3.A.2.2 in Appendix 3.A.2 shows the assignment procedure for the city of Bonn. Using snow grids, average

[7] Verzeichnis der Krankenhäuser und Vorsorge- oder Rehabilitationseinrichtungen in Deutschland

days of snow are calculated for each hospital catchment area. Grids are weighted by their share that is within a catchment area, assuming that the distribution within a grid is constant. In order to account for heterogeneity in catchment areas, grids are additionally weighted by the population density shown in Figure 3.A.2.3 in Appendix 3.A.2. The population weighting factor is a continuous variable that equals zero if the population is zero and one in case of population maximum. The resulting average of those weighted grids within a catchment area is used as the yearly hospital's yearly exposure to days of snow.

Descriptive statistics. Table 3.1 displays the descriptive statistics for the main variables used in the regression. Panel A shows substantial variation in the characteristics of hospitals. The number of beds ranges from 4 to 3,001 and the number of in-patient cases from 77 to 198,452, revealing that the definition of a hospital is independent of its size but rather a legal concept based on permanent availability and equipment. Non-profit and public hospitals account for 43 and 40 percent in the data set. About 17 percent of the general hospitals in the data set are private. However, private hospitals in Germany are obliged to provide the same health services to the same conditions as non-private ones.[8] Between 2006 and 2016, 9 percent of the hospitals were involved in mergers but only 2 percent closed. Panel A in Figure 3.A.1.3 in Appendix 3.A.1 shows a constant increase in the share of private, merged and closed hospitals over time.

Panel B in Table 3.1 lists indicators related to the quantity of human resources and special equipment such as CT or MRT scanners. The annual full-time equivalent staff of physicians and nurses is around 64 and 170, respectively. The average number of special equipments is around 80. This reflects the structure of the German hospital sector, characterized by a large number of relatively small hospitals. Panel C shows hospital output factors. The average number of yearly medical treatment volume is 15,434, ranging from 84 to 197,987. The treatment range – expressed by the number of different medical services categories such as transplantations – is 87 on average. Panels B and C in Figure 3.A.1.3 in Appendix 3.A.1 show that input and output factors increase over time. However, increases in output factors are much stronger.

Panel D in Table 3.1 shows a substantial variation in the base rate deviation between federal and hospital base rates in 2004, ranging from -3,534 € to 1,666 €. This is unsurprising given the strong variation in hospital characteristics, which makes it more difficult for hospitals to replicate average costs. Strong variation also applies to the instrument in Panel E, ranging between -32 and 82 population weighed days of snow, whereby the mean is around 24 days. It reveals a sizable change in exposure to days of snow across hospitals, which

[8] Three types of hospital ownership are defined by German Law: public, owned by the state a federal state or a municipality; non-profit, owned by non-profit associations like the Red Cross or institutions of the church and private hospitals, which primarily aim to make profit by individuals or legal entities (Wissenschaftliche Dienste, 2014)

can be attributed to the mild and changeable weather conditions in Germany (see Panel E) and the heterogeneous distribution of hospitals.

Municipality characteristics of hospital locations in Panel F are provided by the Federal statistical Office, confirming the heterogeneous distribution of hospitals across Germany. For example, while the smallest municipality has 4,000 inhabitants, the largest has 3.5 million inhabitants.

Table 3.1: Descriptive statistics of hospital characteristics

	Mean	SD	min	max	N
A. Hospital structure					
Non-profit	0.43	0.50	0	1	8626
Public	0.40	0.49	0	1	8628
Private	0.17	0.38	0	1	8626
Merged	0.09	0.15	0	1	8626
Closed	0.02	0.05	0	1	8626
Number of Beds	375.46	312.68	4	2917	8626
Inpatients	15666.47	14257.88	77	198452	8626
Base rate in € in 2004	2674.98	362.121	871	7238	1123
B. Human capital and equipment					
Physicians	63.55	135.49	5	3952	8626
Nurses	170.25	287.96	4	4395	8626
Special equipment	80.41	92.90	1	221	8626
C. Services					
Treatment range	87.47	88.72	5	232	8626
Treatment volume/100	154.63	144.38	84	197.98	8626
D. Base rate deviation					
$(br_{ft_{2004}} - br_{it_{2004}})$	21.29	382.56	-3533.97	1666.29	1123
E. Instrument					
$(\Delta snow_{ft_{2003},t_{2004}} - \Delta snow_{it_{2003},t_{2004}})$	23.84	30.34	-32.01	81.68	1123
F. Municipality characteristics					
Inhabitants/1000	256.98	628.47	0.40	3574.83	8626
Employed/1000	112.28	250.20	0.00	1367.68	8626
Share male < 30 years	0.32	0.03	0.23	0.41	8626
Share male 30 - 64 years	0.50	0.02	0.43	0.55	8626
Share male > 64 years	0.18	0.02	0.13	0.27	8626
Share female < 30 years	0.29	0.03	0.20	0.39	8626
Share female 30 - 64 years	0.47	0.02	0.41	0.53	8626
Share female > 64 years	0.23	0.03	0.16	0.34	8626
E. Weather characteristics					
Mean temperature (°C)	9.53	1.41	-5.07	12.34	8626
Mean precipitation in mm/m^2	2.15	0.51	0.78	5.81	8626
Mean Wind speed (m/ss)	3.32	0.93	1.12	11.09	8626

Notes:This table displays the descriptive statistics for the most important variables. Observation in Panels D and E are based on the number of general hospitals in 2004. The data underlying the statistics in Panel E is calculated for hospital catchment areas based on driving time. Panel F is based on the municipality in which a hospital is located in.

3.3 Empirical Strategy

In this section, it is presented the empirical model and the identifying assumption that are necessary to interpret the estimated relationship as causal. I present two alternative identification strategies: first, I apply a selection-on-observable strategy; and second, I apply an instrumental variable approach that exploits idiosyncratic days of snow in a critical time window. For the instrument, I focus on the population weighted deviation in the change in the number of days of snow between 2003 and 2004 between single hospital catchment areas and the remaining catchment areas in a federal state.

3.3.1 Basic model

The staggered convergence phase of the base rate from 2005 until 2010 motivates a continuous treatment setting with the following empirical model, which I apply to general hospitals in Germany over the 2006-2016 period.

The basic model reads:

$$y_{ict} = \alpha + \beta \left(br_{ft_{2004}} - br_{it_{2004}}\right) \times \delta_t + \boldsymbol{X'_{ict}}\gamma + \delta_c + \tau_t + \varepsilon_{ict}, \tag{3.2}$$

where y_{ict} indicates the outcome in year t measured at hospital i located in city c. The main term of interest $(br_{ft_{2004}} - br_{it_{2004}}) \times \delta_t$ is the deviation of the base rate $br_{it_{2004}}$ of a hospital i in 2004 from the base rate $br_{ft_{2004}}$ of a federal state f in 2004 multiplied by the year indicators δ_t. I include the year indicators δ_t to mimic the step-wise harmonization of the base rate, expecting increasing effect sizes over time. Thus, β reveals the average year-specific changes in a hospital's resources between 2006 and 2016 if the hospital base rate deviates from the federal base rate in 2004. In comparison to a long-difference regression with only two observation periods, this model allows me to flexible compare the results with the related literature, even if observation periods differ. The vector $\boldsymbol{X'_{ict}}\gamma$ controls for a number of time-varying characteristics at the level of hospitals and for city population characteristics including population size, employment as well as the city population's composition by age groups and gender. Additionally, I include a set of weather controls measured at the closest weather monitor to the hospital (see Table 3.1 for details). In order to capture time-varying characteristics across municipalities and federal states, I also include municipality and federal state-specific linear time trends. Finally, municipality fixed effects δ_c capture any time-invariant municipality characteristics while year fixed effects τ_t control for any time-specific effects that are uniform across all hospitals. The error term ε_{ict} summarizes all determinants of the proportion of outcomes not captured by the regression. Standard errors are clustered at the county level.

Non-linearity. Given that treatment prices increase for some hospitals and decrease for others, non-linear effects of price changes are plausible if the effects of price increases and decreases are asymmetric. I test this by comparing the effects on outcomes for different ranges of price shocks using a spline regression.

Variation over time. As the base rate convergence fades out in 2010 while the observation period lasts until 2016, I provide evidence on the heterogeneous effect over time. For this purpose, I estimate a difference-in-difference specification whereby I interact eight mutually exclusive year dummies between 2006 and 2016 with the base rate deviation from 2004.[9]

$$y_{ict} = \sum_{t=1}^{8} \delta_t \times (br_{ft_{2004}} - br_{it_{2004}}) + \boldsymbol{X'_{ict}}\gamma + \delta_c + \tau_t + \varepsilon_{ict}. \tag{3.3}$$

The coefficient δ_t measures the effect of an increase in the difference between the hospital and federal base rate in 2004 on hospital resources for the eight year dummies $t = 1, \ldots, 8$, controlling for the same characteristics as before. In this analysis, I do not focus on the organizational structure because hospitals do not demerge or re-open.

3.3.2 Instrument Variable Strategy

Hospital base rates are potentially endogenous in 2004 if they are correlated with structural hospital characteristics, which would bias results towards zero. I exploit external variation in hospital base rates by increases in treatments using idiosyncratic deviations in weather conditions to which hospitals are exposed. In the economic literature, precipitation has been established as a valid instrument (e.g. Almond et al. (2009), Maccini and Yang (2009)). Days of snow are found in the epidemiological literature to have a strong positive impact on the number of cost intensive hospital admissions. Especially for fractures and other types of injuries caused by falling on slippery ground (Ralis, 1981; Björnstig et al., 1997) as well as cardiovascular events due to physically-demanding activities like snow shoveling, which increases blood pressure in a situation where cold temperatures narrow blood vessels (Franklin et al., 1995; Auger et al., 2017). The fact that Germany is located in a moderate climate amplifies the effect of snowfall on hospital admissions. Due to a high variation of weather conditions, it is difficult for individuals to avoid accidents by adjusting behavior (Eisenberg and Warner, 2005). Therefore, I use the deviation in the change between 2003 and 2004 in the population weighted number of days of snow between single hospitals catchment areas and the remaining catchment areas in the same federal state to instrument hospital base rates.

[9] Hospital Quality reports were not reported in 2007, 2008 and 2011

The idea behind this instrument is motivated by a peculiarity of the G-DRG-System convergence phase. It started in 2004 with a budget-neutral year in which hospitals negotiated their budget with health insurance companies based on the old system at the end of 2003. Higher levels of days of snow in 2003 improved hospitals' bargaining position as it enabled them to draw attention to the increasing need for treatments. The negotiated budget was then divided by the case mix of 2004, which is the number of in-patient cases per year weighted by their severeness. This was undertaken to calculate a hospital-specific base rate, representing the starting point of the convergence phase in 2005. However, if hospitals were exposed to higher levels of snow in 2004 compared with 2003, case mixes increased due to increasingly more severe admissions. As a result, hospital base rates decreased. Furthermore, federal base rates were calculated based on the accumulated hospital budgets in a federal state and divided by the accumulated case mixes in a federal state. Thus, deviations in days of snow between 2003 and 2004 affected the federal state base rate similar to hospital base rates. A single hospital would face lower treatment prices over the convergence period if days of snow in its own catchment area in 2004 occurred more frequently than in 2003 or if days of snow in the remaining catchment areas in the federal state occurred more frequently than in 2004. I capture this mechanism with the following approach.

The first stage regresses the deviation of the hospital from the federal base rate in 2004 on the instrument, namely the population-weighted deviation of changes in days of snow between 2003 and 2004 conditional on the average days of snow.

$$(br_{ft_{2004}} - br_{it_{2004}}) = \lambda_0 + \lambda_1(\Delta snow_{ft_{2003},t_{2004}} - \Delta snow_{it_{2003},t_{2004}}) + \boldsymbol{X'_{ict_{2004}}}\boldsymbol{\kappa} + \delta_c + \varepsilon_{ic}, \quad (3.4)$$

I measure the instrument $\lambda_1(\Delta snow_{ft_{2003},t_{2004}} - \Delta snow_{it_{2003},t_{2004}})$ as the deviation of the change in the number of days of snow between 2003 and 2004 $(\Delta snow_{it_{2003},t_{2004}})$ in a catchment area of hospital i from the change of days of snow between 2003 and 2004 in the remaining catchment areas $(\Delta snow_{ft_{2003},t_{2004}})$ in a federal state f, additionally controlling for hospital and municipality characteristics in 2004 and city fixed effects δ_c[10]. Furthermore, I control for average days of snow in hospital catchment areas before 2003. The instrument predicts a hospital's base rate deviation from the federal base rate based on the abnormal change in the number of days of snow between 2003 and 2004 between a hospital's catchment area and the remaining catchment areas in a federal state.

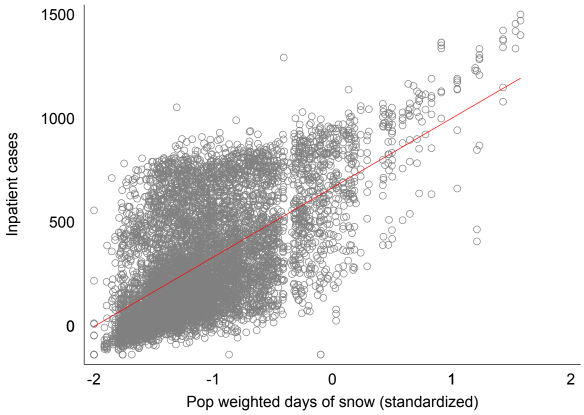
In order to demonstrate the relevance of the instrument, I show the effect of population-

[10]Figure 3.A.2.4 in Appendix 3.A.2 shows that days of snow in 2003 and 2004 are strongly correlated with the size of the hospital. However, these differences disappear when adding municipality fixed effects.

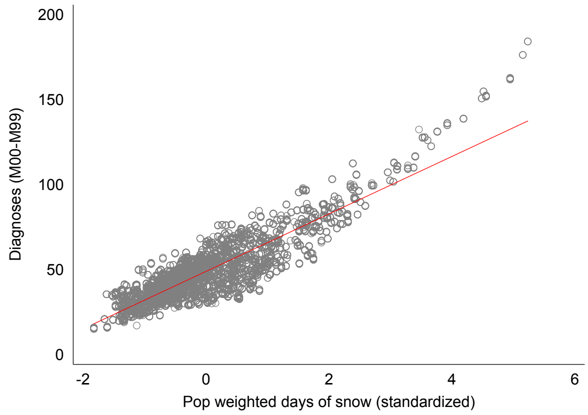
weighted days of snow in hospital catchment areas on in-patient cases in general and for cases that should be especially affected by days of snow. Panel A in Figure 3.2 displays the correlation between the instrument and in-patient cases between 2006 and 2016 after controlling for hospital and municipality characteristics and municipality and year fixed effects. As I do not have data on diagnoses as prior to 2006, I assume that days of snow affect in-patient cases after 2006 in the same manner as before 2006. A one-standard-deviation increase in population-weighted days of snow increases the number of in-patient cases by around 400 cases or 2.5 percent at the mean. Using the same specification, the same picture is true for diagnoses regarding the musculoskeletal system, connective tissue (Panel B) and the circulatory system (Panel C), which should be most sensitive to snow. For example, a one-standard-deviation increase in population-weighted days of snow increases the number of musculoskeletal system and connective tissue diagnoses by around 6 percent at the mean. Thus, days of snow in hospital catchment areas have a direct impact on the number of in-patient cases and should affect hospitals' base rate.

Panel D in Figure 3.2 shows the first-stage correlation of Equation 3.4. The correlation is strong and has the expected positive sign, resulting from an increase in the difference in the change of exposure to days of snow between the single hospital catchment area and the remaining catchment areas in a federal state between 2003 and 2004. A one-standard-deviation increase in the instrument increases the base rate deviation by around 10 percent of a standard deviation. Detailed information about the F-statistic and the first stage results is shown in Table 3.3.

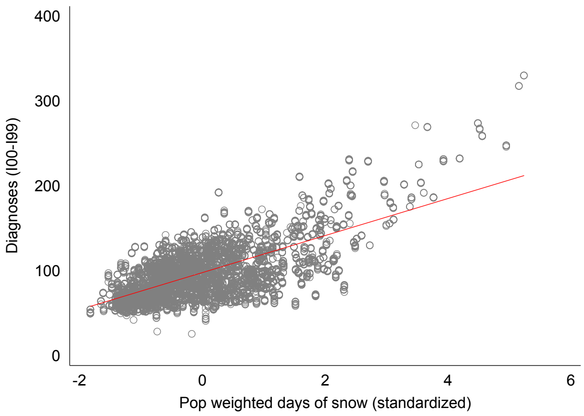
As a robustness check, I replicate the IV approach with non-exclusive catchment areas based on 10 minutes driving time around a hospital location. Table 3.A.3.1 in Appendix 3.A.3 shows similar first-stage results and comparable F-statistics compared to Table 3.3.



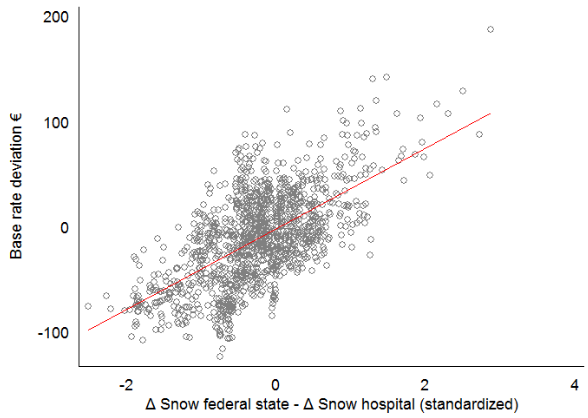
(a) In-patient cases



(b) Musculoskeletal system and connective tissue diagnoses



(c) Diseases of the circulatory system



(d) Hospital base rate

Figure 3.2: First-stage correlation

Notes: Panels A-C) display the correlation between hospital admissions and days of snow in hospital catchment areas between 2006 and 2016, conditional on hospital and municipality characteristics, municipality and year fixed effects and average snowfall between 2000 and 2003. Panel D) displays the first-stage correlation between the standardized instrument $(\Delta snow_{ft_{2003},t_{2004}} - \Delta snow_{it_{2003},t_{2004}})$ *and the base rate deviation in 2004 by controlling for the hospital and municipality characteristics, municipality and year fixed effects and average snowfall between 2000 and 2003.*

Exclusion restriction. To represent a valid IV, days of snow have to satisfy the exclusion restriction that changes between 2003 and 2004 in the deviation of days between a hospital's catchment area and the remaining catchment areas in a federal state – conditional on the yearly mean of days of snow and other controls – only affects hospital resources through its effect on base rates. Variation in days of snow could simultaneously affect the base rate and resources allocations; for example, if small hospitals sort into places with more days of snow. However, this would not violate the exclusion restriction because I control for average days of snow in catchment areas and municipality fixed effects.

In order to confirm the exclusion restriction, I perform balancing and falsification tests. In a balancing test, I regress the number of hospital beds in 2004 on the instrument. A significant coefficient would indicate that the instrument is invalid due to its correlation with the error term of Equation 3.2. Figure 3.A.2.4 in Appendix 3.A.2 shows that once controlling for municipality fixed effects, the instrument does not predict the number of beds. This finding is consistent with the assumption that the instrument is as good as randomly assigned. I further perform falsification tests by using the same instrument as in Equation 3.4 but now based on the change of days of snow between 2004 and 2005 as well as 2005 and 2006 to test the validity of the instrument. Finding high F-statistics after 2004 would indicate that the instrument picks up snow patterns that are correlated with hospital resources, thus violating the exclusion restriction. Column (1) in Table 3.A.3.2 in Appendix 3.A.3 replicates the F-statistic of Table 3.3 and Columns (2) and (3) show the F-statistics of the falsification tests. The F-statistics in Columns (2) and (3) of Table 3.A.3.2 in Appendix 3.A.3 are weak. Based on the same idea, I perform a falsification test in Table 3.A.3.3 in Appendix 3.A.3 where I focus on special hospitals that treat snow-related admissions less often. Column (1) replicates the F-statistic of Table 3.3, whereas Column (2) shows the F-statistic for the sample of special hospitals. As expected, the F-statistic in Column (2) is weak. This is again consistent with random sampling variation around the true effect of zero.

Thus, the exclusion restriction would only be violated if changes in the number of days of snow between 2003 and 2004 affected hospital resources through a channel other than the base rate. I assume that this is implausible because variation in days of snow between 2003 and 2004 should not trigger strategic management decisions that affect resources in the following years.

3.4 Results

In this section, I present the estimation results for the effect of treatment price shocks on hospital resource reallocation. I first show OLS estimates to compare the results with the existing literature and discuss extensions such as non-linear effects and the variation of effect sizes over time. In order to address concerns about endogeneity, I show IV estimates.

3.4.1 OLS Results

Table 3.2 displays the OLS results. Each coefficient is the result of a separate regression of the outcome listed on the left on the treatment price shocks and the controls listed at the bottom. The outcomes in Panels A and B are standardized to a mean of zero and a standard deviation of one. As the results in Panel C represent percentages based on binary indicators, I do not standardize these outcomes. A coefficient of $\hat{\beta} = 0.01$ in Panel A and B means that an increase in prices by 1 € over the convergence phase – which equals a 1 €

deviation between the federal and the hospital base rate in 2004 – is associated with an increase in the respective hospital resources by 1 percent of a standard deviation per year between 2006 and 2016. I discuss the effect size relative to an increase in prices by one standard deviation (sd(€) = 382,56), to facilitate the interpretation. In Panel C, the base rate deviation is multiplied by 100.

Column (1) reports the OLS estimates from binary regressions without controls. All coefficients are small and weakly significant. After controlling for year and municipality fixed effects in Column (2), the point estimates become larger in absolute terms. I include municipality fixed effects to reduce the potential bias from residential sorting. As shown in Figure 3.A.1.1 in Appendix 3.A.1, larger hospitals have higher base rates, although they are more frequently located in urban areas. Thus, controlling for municipality fixed effects captures confounding differences between hospitals with different sizes. Year fixed effects capture general trends in the hospital sector like demographic or technical changes. Controlling for hospital characteristics in Column (3) increases point estimates while controlling for weather characteristics in Column (4) leave the results rather unchanged. The coefficients between Columns (4) and (5) become larger again when controlling for linear municipality and federal state time trends. Linear municipality time trends capture the urbanization of hospital supply, which is driven by migration towards cities and the political will to relocate more medical services to the city. However, the hospitals that face an increase in demand due to urbanization are those that experience base rate increases less often due to their larger size. This confounding effect is captured by linear municipality time trends.

If I include the full set of controls in Panel A, a one-standard-deviation increase in treatment prices is associated with a yearly increase in the number of physicians by 0.6 percent[11] and the number of nurses by 0.7 percent of a standard deviation. Investments in special equipment are unaffected. This can be explained by the fact that G-DRGs do not cover capital investments because hospitals receive additional funds for investments from municipalities or federal states. Panel B shows an increase in the treatment range if treatment prices increase. A one-standard-deviation increase in treatment prices is associated with a yearly increase in the range of treatments by 1.2 percent of a standard deviation. The treatment volume decreases by 1.1 percent of a standard deviation per year. Panel C shows that an increase in the treatment prices by 100 € reduces the probability of privatization by 0.3 percentage points per year and the probability of being merged by 0.2 percentage points per year. The effect sizes on closures are relatively large. However, this is based on a very small sample as only 2 percent of all hospitals were closed in the observation period. Table 3.A.3.4 in Appendix 3.A.3 displays the adjusted R^2 for the OLS regressions in Table 3.2.

[11] For example, in Column (5) of Table 3.2 an increase in the base rate deviation of one standard deviation ($= 383$) translates into an effect size for a standard deviation of physicians ($= 135$) of $0.002 \times 383/135 = 0.006$, i.e., 0.6 percent.

Table 3.2: OLS results: the effect of treatment prices on hospital resources

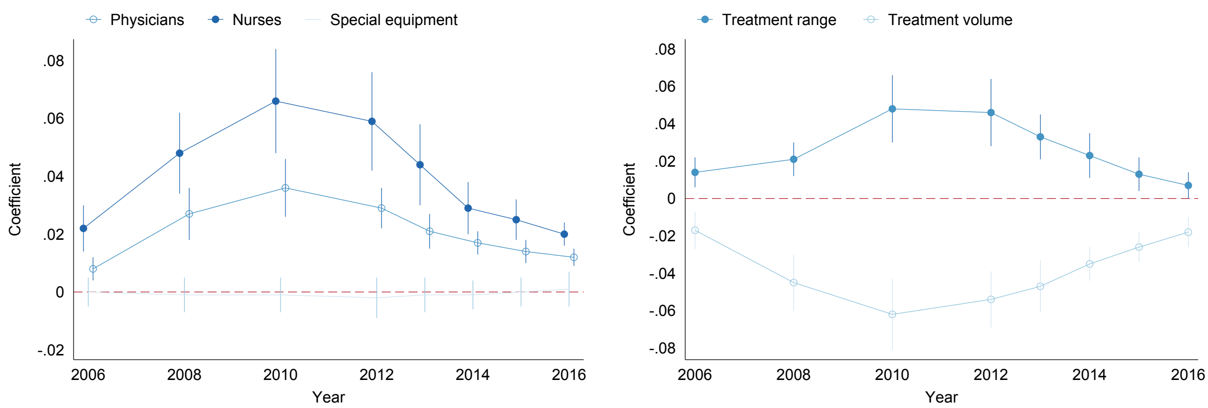
	(1)	(2)	(3)	(4)	(5)
A. Input					
Physicians	0.001*	0.002**	0.002*	0.002**	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Nurses	0.001	0.002**	0.004**	0.004**	0.005***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Special equipment	-0.000	-0.001	-0.002	-0.001	-0.001
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
B. Output					
Treatment range	0.001*	0.002**	0.002***	0.002***	0.003***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
Treatment volume	-0.001	-0.002**	-0.003***	-0.003***	-0.004***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
C. Structure					
Private	0.000	-0.002**	-0.003**	-0.003**	-0.003**
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Merged	-0.001	-0.001**	-0.001**	-0.002**	-0.002**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Closed	-0.000	-0.001*	-0.001*	-0.001*	-0.001*
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
N	8626	8626	8626	8626	8626
Controls:					
Year FE	No	Yes	Yes	Yes	Yes
Municipality FE	No	Yes	Yes	Yes	Yes
Hospital characteristics	No	No	Yes	Yes	Yes
Weather characteristics	No	No	No	Yes	Yes
Municipality characteristics	No	No	No	No	Yes
Linear time trends	No	No	No	No	Yes

Notes: This table displays the OLS results. Each coefficient is the result of a separate OLS regression of the outcome listed on the left on the base rate in €, controlling for the variables indicated below. In Columns (1)-(5) of panel A and B, the hospital resources have been standardized. In Panel C, outcomes have not been standardized. However, the base rate deviation is multiplied by 100. Standard errors are displayed in parentheses and clustered at the county level. Significance level: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Non-linear Effects. In Table 3.A.3.5 in Appendix 3.A.3, I test for a non-linear relationship. Based on spline regression of hospitals above and below the average federal base rate in

2004, I find no evidence of non-linear effects. Thus, the transition from a base rate decrease to an increase seems to be linear. Put differently, an increase in the base rate deviation between hospitals and federal base rates in 2004 has the same effect for hospitals above and below the federal base rate. This leads to a polarization of resources between hospitals that faced treatment price increases and hospitals that faced treatment price reductions.

Variation over time - OLS. Figure 3.3 plots the year-specific effects of treatment price increases on input (Panel A) and output factors (Panel B). From 2006 onwards, treatment price shocks already affect input and output factors. Until 2010, hospitals gradually adjust their resources. If the treatment price increase, the number of physicians, nurses and the treatment range increases, whereas the amount of special equipment remains rather constant and the treatment volume decreases. In all resource dimensions, the peak seems to be reached by the end of the convergence phase in 2010. For example, in 2010 the number of physicians and nurses are 10.2 percent and 8.8 percent of a standard deviation higher for a one-standard-deviation increase in treatment prices. Correspondingly, the treatment range in 2010 is 19.3 percent of a standard deviation higher while the treatment volume is 16 percent of a standard deviation lower. Input and output factors tend to converge to the pre-reform levels after 2010. However, differences still remain until 2016, indicating a persistent pattern of treatment price shocks.



(a) Input

(b) Output

Figure 3.3: Variation over time - OLS

Notes: These graphs plot the estimated OLS coefficients and the 95 percent confidence intervals for the year-specific effect of the base rate deviation in 2004 on hospital resources. The estimates are based on Equation 3.3. Standard errors are clustered at the county level.

3.4.2 IV Results

In Table 3.3, I present the instrumental variable estimates from regressions with all controls and year and municipality fixed effects and linear time trends. Column (1) reproduces the OLS results from Column (5) in Table 3.2. Column (3) report the first-stage coefficients and F-statistics of the IV estimation based on 1,123 observations in 2006. The reduced-form estimates in Column (2) have the expected signs. A higher deviation of days of snow between a single hospital catchment area and the remaining hospital catchment areas within a federal state means a higher deviation between the federal and the hospital base rate, which has a positive effect on the number of physicians, nurses and the diversity of treatments, and a negative impact on treatments, privatization, mergers and closures. Seven out of eight coefficients are statistically significant at the 5 percent-level. While the magnitude of the reduced-form coefficient is not straightforward to interpret, the statistical significance is important for the interpretation of the second-stage results. It rules out the possibility that the second-stage results in Column (3) are a statistical artifact stemming from sampling variation in the first stage.

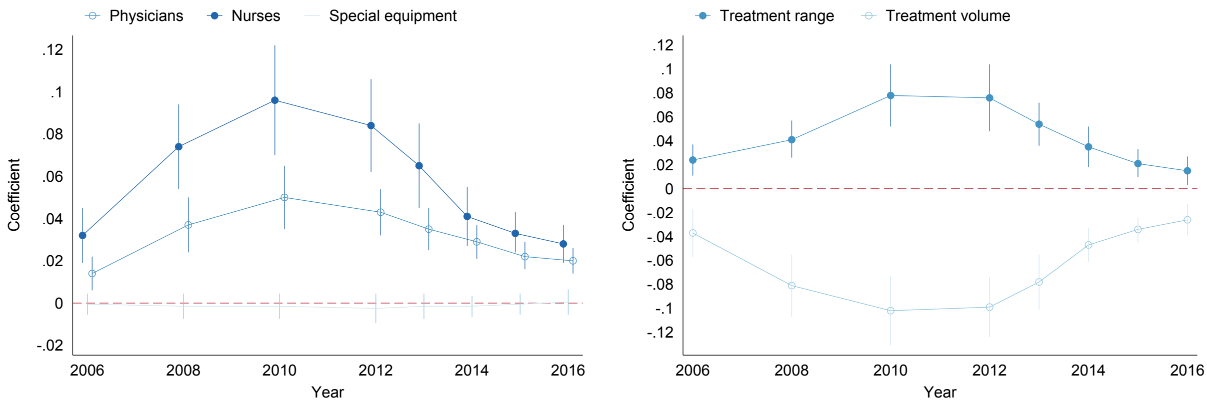
The second-stage results are larger than the OLS results. A one-standard-deviation increase in treatment prices is associated with a yearly increase in the number of physicians by 1.1 percent and the number of nurses increases by 0.9 percent of a standard deviation. Panel B shows also stronger magnitudes of second-stage results compared with the OLS results. A one-standard-deviation increase in treatment prices is associated with a yearly increase in the treatment range by 2.5 percent of a standard deviation, while the treatment volume decreases by 2.4 percent of a standard deviation. Panel C shows that an increase in treatment prices of 100 € reduces the probability of privatization by 0.7 percentage points per year and the probability of being merged by 0.4 percentage points per year.

Table 3.3: IV results: the effect of treatment prices on hospital resources

	OLS (1)	Reduced form (2)	IV-2SLS (3)
A. Input			
Physicians	0.002***	0.150***	0.004**
	(0.000)	(0.035)	(0.002)
Nurses	0.005***	0.200**	0.007**
	(0.001)	(0.081)	(0.003)
Special equipment	-0.001	-0.180	-0.002
	(0.000)	(0.130)	(0.002)
B. Output			
Treatment range	0.003***	0.130***	0.006***
	(0.001)	(0.035)	(0.002)
Treatment volume	-0.004***	-0.101***	-0.009***
	(0.001)	(0.031)	(0.003)
C. Structure			
Private	-0.003**	-0.219***	-0.007*
	(0.001)	(0.049)	(0.004)
Merged	-0.002**	-0.199***	-0.004*
	(0.001)	(0.053)	(0.002)
Closed	-0.001*	-0.143**	-0.001
	(0.000)	(0.051)	(0.000)
First-stage:			
Snow instrument			50.922***
			(10.006)
F-statistic			47.238
N	8626	8626	
Controls:			
Year FE	Yes	Yes	
Municipality FE	Yes	Yes	
Hospital characteristics	Yes	Yes	
Weather characteristics	Yes	Yes	
Municipality characteristics	Yes	Yes	
Linear time trends	Yes	Yes	

Notes: This table displays the IV results. Column (1) reproduces the OLS results from Column (5) in Table 3.2. Column (2) reports the reduced-form coefficients of separate regressions of the outcomes on the instrument and controls. The main IV results are displayed in Column (3). Column (3) additionally reports the first-stage coefficients of the base rate regressed on the instrument and all controls mentioned in section 3.3.1 and the corresponding F-statistics. Standard errors are displayed in parentheses and clustered at the county level. Significance level: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Variation over time - IV. Figure 3.4 plots the year-specific effects of treatment price increases on input (Panel A) and output factors (Panel B), based on split sample 2SLS per year. The magnitudes are larger than in Figure 3.3, while the general picture remains rather constant. From 2006 onwards, treatment price shocks already affect input and output factors. Until 2010, hospitals gradually adjust their resources. The peak seems to be reached by the end of the convergence phase in 2010. For example, in 2010 the number of physicians and nurses are 12 percent and 12.8 percent of a standard deviation higher for a one-standard-deviation increase in treatment prices. Correspondingly, the treatment range in 2010 is 35 percent of a standard deviation higher, while the treatment volume is 27 percent of a standard deviation lower. Subsequently, input and output factors tend to converge to the pre-reform levels. As in the case of the previous OLS results, differences remain until 2016, confirming a persistent pattern of treatment price shocks.



(a) Input (b) Output

Figure 3.4: Variation over time - IV

Notes:Notes: These graphs plot the estimated IV-coefficients and the 95 percent confidence intervals for the year-specific effect of the base rate deviation in 2004 on hospital resources. The estimate is based on Equation 3.3. Standard errors are clustered at the county level.

Interpretation of IV estimator. There are three potential reasons why the IV estimator – given that the instrument is valid and strong – produces different results from an OLS estimator. First, the difference can be explained by heterogeneous treatment effects. If treatment effects are not constant across hospitals, the IV estimator identifies a local average treatment effect (LATE). With a continuous instrument, this means that the estimator places a stronger weight on hospitals with increasing treatment prices, which are typically small hospitals. In Figures 3.A.3.1 a and 3.A.3.1 b in Appendix 3.A.3, I explore potential sources of heterogeneous treatment effects. I do not find a stronger first-stage relationship for hospitals that experience decreases compared with increases in treatment prices or differences between large and small hospitals. Thus, there is no identification of a LATE. A second reason could be measurement error of the treatment prices. Given that I use official data of hospital base rates, this is rather unlikely. This suggests that OLS estimates are confounded due to unobserved heterogeneity that is not absorbed by controls and fixed effects. The true effect is biased towards zero, which indicates that resource allocations before 2004 correlate with resource allocations in the convergence phase, as described in section 3.2.1.

3.4.3 Discussion of the main results

The estimates presented in sections 3.4.1 and 3.4.2 show that treatment price shocks induced reallocations of hospital resources. These price shocks can be substantial. For an average hospital, a one standard deviation increase in prices translates into a budget increase of 5.9 million Euros between 2005 and 2010.[12] I show that treatment price shocks can have persistent effects on hospital resource reallocations even when treatment price shocks vanish after 2010. Persistent pattern in price shocks have been observed for several sectors where firms pass on price increases to consumers faster than decreases (Peltzman, 2000). This is the first paper to show persistent pattern in the reallocation of hospital resources if prices change.

To assess the magnitude of the results, it is useful to compare them with results obtained in other studies exploiting treatment price shocks in hospitals. The effect sizes and even the sign of the magnitude vary across studies. According to the estimates of the federal budgeting for medical reimbursement in the United States, a one-percent decrease in treatment prices increases the treatment volume by around 0.3-0.5 percent. By contrast, Clemens and Gottlieb (2014) show that higher treatment prices increase the treatment volume, whereas Dafny (2005) finds no effect. Using the G-DRG-Reform, Salm and Wübker (2015, 2018) find results that are smaller compared to the results in this chapter but they do find linear effects as well. For example, a one-percent increase in treatment prices decreases the treatment volume by 0.14 percent until 2009, the number of nurses by 0.4 percent and the number of physicians by around 0.2 percent until 2010. OLS estimations in this chapter, which include the post convergence period after 2010 – thus representing a lower-bound estimate – show that an increase in treatment prices by one percent of the mean (mean(€) = 27) decreases the treatment volume by 0.23 percent[13] until 2009, the number of physician by 0.34 percent and the number of nurses by 0.38 percent until 2010. Results are even larger in this chapter for 2010 when interacting mutually exclusive year dummies with the base

[12] One standard deviation in the deviation between the federal and the hospital base rate in 2004 equals around 383 €. The average treatment volume of a hospital is around 15,400 cases per year.

[13] For example, in Column (5) of Table 3.2 an increase in the base rate deviation of one percent of the mean ($= 27$) translates into an effect size for the mean of treatments ($= 144.38$) from 2006 to 2009 of $0.004 \times 27/144.38 \times 3 = 0.0025$, i.e., 0.25 percent.

rate deviation from 2004, thereby analyzing heterogeneous effects over time. Furthermore, estimations in this chapter, which rely on plausibly exogenous variation in treatment price shocks confirm larger effect sizes. IV results – which include the post convergence period after 2010 – show that the number of physicians actually increased by 0.85 percent, the number of nurses by 0.7 percent and the treatment volume by 0.62 percent on average per year. Again, effect sizes in 2010 are much larger when analyzing heterogeneous effects over time.

3.4.4 Heterogeneous Effects

In Table 3.4, I test whether the impact of treatment price shocks differs between major hospital characteristics, by focusing on pre-treatment ownership, hospital size and urban or rural locations in 2004. The selection of pre-treatment characteristics is based on the information available in the German Hospital Directory. However, the hospital ownership and size are good proxies for the efficiency of a hospitals. Furthermore, hospitals in rural and urban areas have experienced different trends. Hospital supply has been reallocated to cities due to the inefficiency of rural hospitals. In case of treatment price shocks, inefficiency should lead to heterogeneous effects.

For each set of groups, I estimate split sample 2SLS. In all regressions, I control for the same characteristics as in Column (5) of Table 3.2. Comparing Columns (1) and (2) shows different treatment effects between hospitals that were public or non-public in 2004. With the exception of special equipment, input and output factors of non-public hospitals are more strongly affected compared with public hospitals. This means that public hospitals react less to treatment price decreases given the finding of linearity in price changes in Table 3.A.3.5 in Appendix 3.A.3. In case of input factors, this finding is in line with more flexible employment in the non-public sector. Smaller impacts on the treatment range and the treatment volume in Panel B could indicate more supplier-induced demand among private hospitals. Finding stronger effects for non-public hospitals could also be explained by the fact that municipalities – which are the owners of public hospitals – might subsidize their hospitals.

In Columns (3) and (4), I compare hospitals with above- and below-median numbers of beds[14] in 2004. Again, a rather general effect occurs, whereby larger hospitals tend to react more strongly to treatment price shocks compared with smaller hospitals. One potential explanation is the general financial pressure among large hospitals. They faced difficulties in replicating the average treatment costs of a federal state expressed by higher base rates in 2004 shown in Figure 3.A.1.1 in Appendix 3.A.1. Furthermore, larger hospitals can reduce resources more easily due to economies of scale. Interestingly, smaller hospitals

[14]Median number of beds = 325

merge sightly less often than larger hospitals if treatment prices increase. One explanation could be that it is much easier for smaller hospitals to merge.

Finally, in Columns (5) and (6), I assess if the effects differ between hospital, which were located in urban versus rural areas[15] in 2004. I find no difference in estimates for input factors between rural and urban areas and some minor differences for output factors. The range and the volume of treatments are stronger affected in case of rural hospitals. Rural hospitals might compensate higher inefficiency with more supplier-induced demand. More importantly, the hospital structure in Panel C seems to play a major role between urban and rural areas. A treatment price increase decreases the probability of mergers much stronger in rural areas. This is in line with policies that favor the consolidation of hospital supply, especially in rural areas.

[15] Municipalities with less than 50,000 inhabitants

Table 3.4: Heterogeneous Effects

	Public	Non-public	(Large)	(Small)	(Urban)	(Rural)
	(1)	(2)	(3)	(4)	(5)	(6)
A. Input						
Physicians	0.003*	0.006*	0.006*	0.002*	0.004**	0.004**
	(0.001)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
Nurses	0.005**	0.009**	0.010**	0.005**	0.008**	0.007**
	(0.002)	(0.004)	(0.005)	(0.002)	(0.004)	(0.003)
Special equipment	-0.002	-0.001	-0.001	-0.001	-0.002	-0.002
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
B. Output						
Treatment range	0.004**	0.007***	0.008**	0.004***	0.004**	0.007***
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
Treatment volume	-0.007***	-0.011***	-0.012***	-0.006***	-0.008***	-0.010***
	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)	(0.003)
C. Structure						
Merged	-0.002*	-0.005**	-0.004**	-0.003*	-0.002*	-0.006**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Closed	-0.001	-0.001	-0.001	-0.002*	-0.001	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	3431	5197	4315	4313	5434	3192
Controls:						
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Weather characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Linear time trends	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each cell represents the effect of the base rate instrument on the outcome listed in the Column and the sample described at the top, from an IV specification that includes our baseline set of controls. Robust standard errors clustered at the county level.

3.5 Conclusion

In this chapter, I have examined the causal effect of general treatment price shocks on hospital resource reallocations in Germany. Starting in 2005, hospital payment for patients was transformed to the G-DRG-System, which led to general idiosyncratic treatment price shocks for individual hospitals until 2010. Some hospitals were exposed to price increases, while others experienced price reductions. However, thus far there is little evidence of the impact of general price shocks on the reallocation of hospital resources.

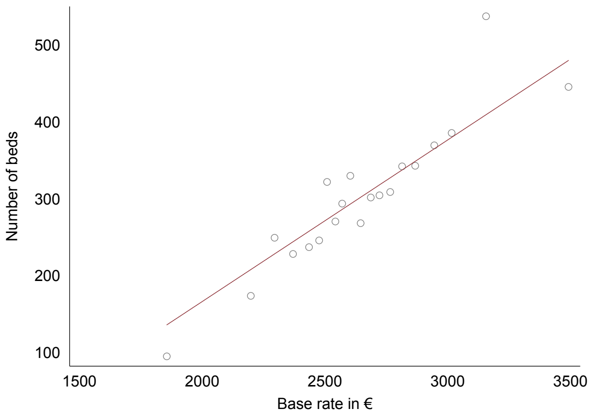
The results of the chapter show that treatment price shocks significantly affected hospital resources. By exploiting data from the universe of German hospitals for the 2006-2016 period, I find that price shocks are positively associated with number nursing staff and physicians and the range of treatments but negatively with the treatment volume. The probability of hospital mergers, closures and privatization is negatively associated with price shocks. Effects are stronger for private and larger hospitals. I show that treatment price shocks can have persistent effects on hospital resource reallocations even when treatment price shocks vanish. Using unique high-resolution satellite data, I implement an novel instrument variable strategy that exploits the exogenous variation in the number of days of snow in hospital catchment areas. A peculiarity of the reform allowed deviations in weather condition at the time of the reform implementation to have an effect on treatment prices in the next five years. IV estimates show that OLS results are biased towards zero for almost all dimensions. Thus, simple OLS regressions would underestimate the true effect due to correlations between structural hospital characteristics and the reallocation of resources after price shocks.

These findings have implications for research and policy. Most research explore reforms where treatment prices change only for a sub-group of patients. This is why there is little evidence about the impact of general treatment price shocks on hospital resources. By exploiting the G-DRG-Reform, I show that hospital reallocate resources due to general treatment price shocks. As this is the first paper showing a persistent pattern in the reallocation of hospital resources if prices change, further research on this topic is necessary. However, this requires well-founded identification strategies. Finding results that are biased towards zero for input and output factors and the organizational structure also points to the importance of unobserved heterogeneity in the context of healthcare policy evaluation.

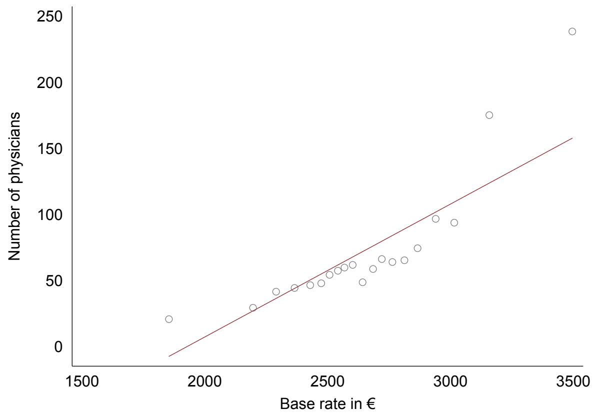
For policy-makers, these results are important for several reasons. The G-DRG-Reform led to a persistent polarization of hospital resources, as some hospitals were exposed to treatment price increases, while others experienced treatment price reductions. If hospitals increase the treatment volume as a response to price decreases by offering unnecessary therapies, it has a negative impact on population well-being and public spending. On the other hand, results show a decrease in the range of treatments if prices decrease. Hospitals might specialize more, thus attracting more patients. From a policy perspective, it is important to evaluate weather such changes in the treatment range jeopardize an adequate nationwide provision of treatments. Furthermore, the results reveal a decrease in the number of nurses and physicians if treatment prices decrease. This could partly explain the nursing shortage in German hospitals. However, since hospitals specialize more they might be able to realize efficiency gains, which justify decreases in input factors without losses of productivity. Another important aspect are changes in the organizational structure, given that many public hospitals became privatized or merged with other hospitals. The findings show that this is at least partly driven by the G-DRG-Reform. However, this can again lead to a lack of services offered in some regions if merged hospitals specialize or if hospitals are taken over by ecclesiastical organizations that do not provide all services due to moral conviction (Filistrucchi and Prüfer, 2018). To have a complete picture, further research is needed on the direct impact of treatment price shocks on healthcare quality.

APPENDIX

3.A.1 Hospital data



(a) Number of beds and hospital base rate



(b) Number of physicians and hospital base rate

Figure 3.A.1.1 : Hospital resources and base rate

Notes: Panel (a) shows a binscatters of the raw correlation between the hospital base rate and the number of beds in 2004. Panel (b) shows the binscatter for the raw correlation between the hospital base rate in 2004 and the number of physicians in 2006 because detailed data of hospital input factors is only available since 2006. Source: Hospital Quality Reports, German Hospital Directory.

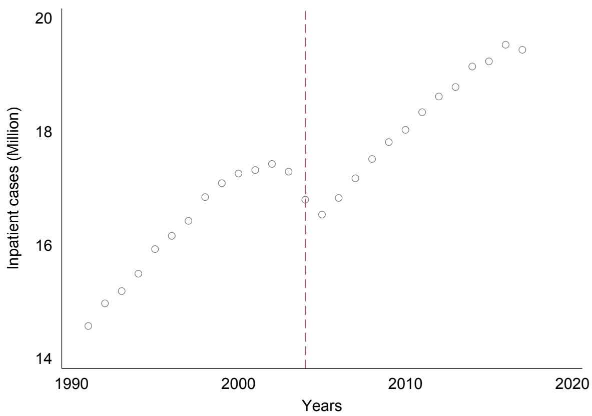
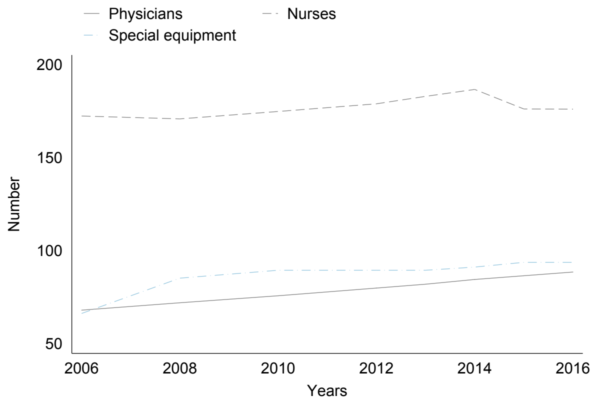
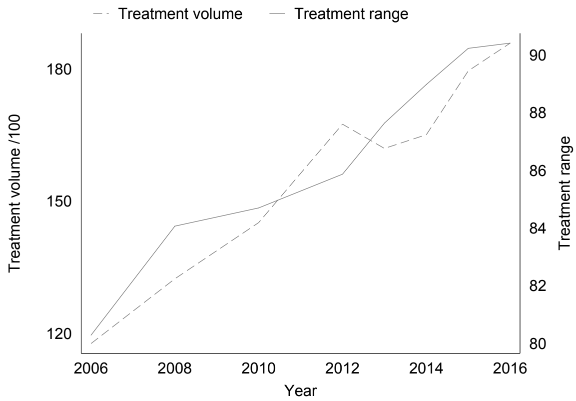


Figure 3.A.1.2 : In-patient cases over time

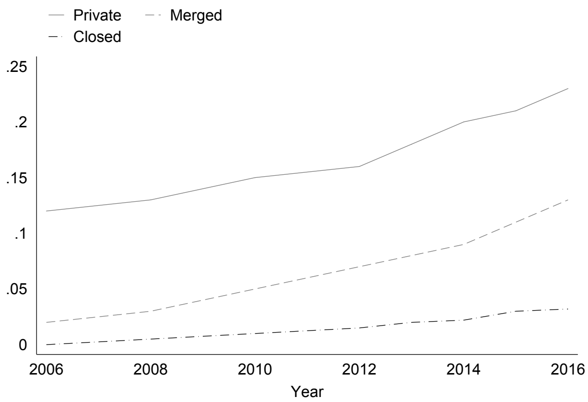
Notes: This figure shows the number of in-patient cases over time. Source: Federal statistical Office.



(a) Input



(b) Output

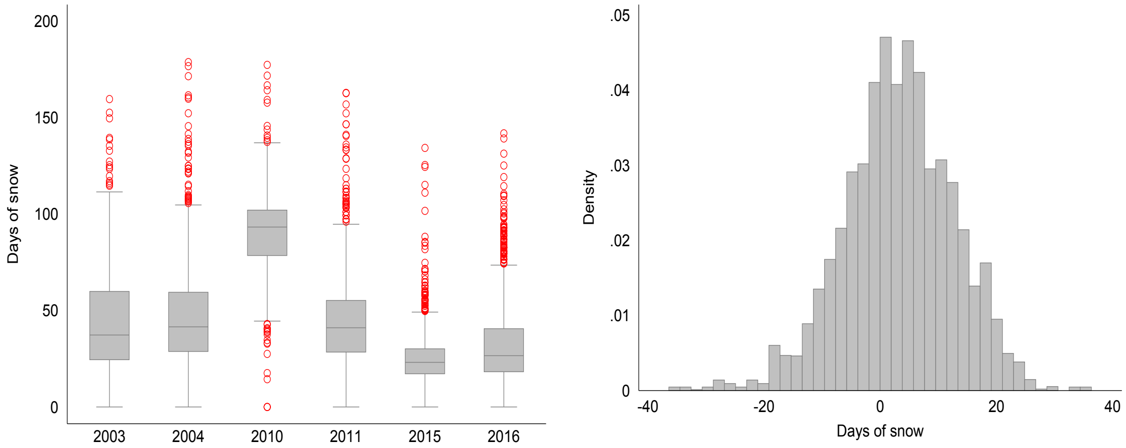


(c) Structure

Figure 3.A.1.3 : Input, output and structure over time

Notes: Panel (a) shows the average number of hospital input factors over time. Panel (a) shows the average number of hospital output factors over time. Panel (c) shows the share of private, closed and merged hospitals over time. Source: Hospital Quality Reports.

3.A.2 Catchment areas and days of snow



(a) Days of snow per year

(b) Deviation in days of snow between 2003 and 2004

Figure 3.A.2.1 : Days of snow

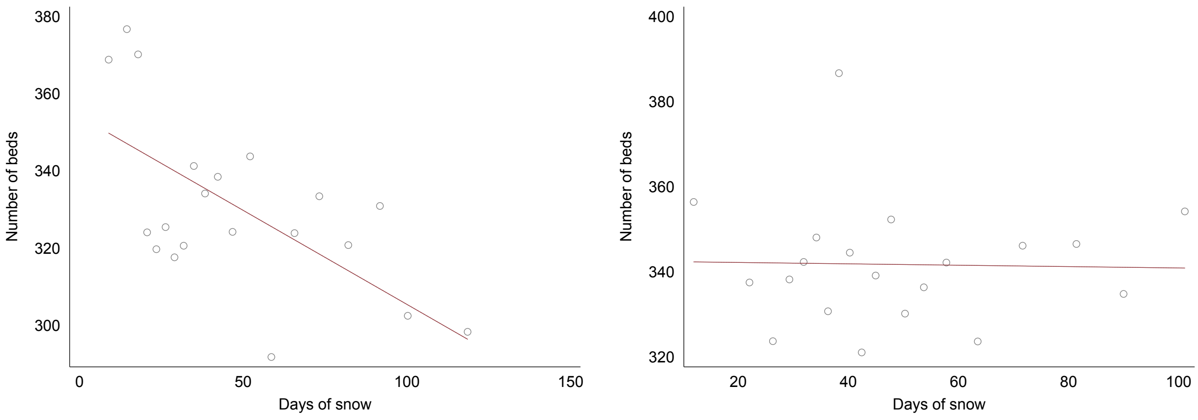
Notes: This graph displays the variation in days of snow over time. Panel (a) displays the variation in days of snow between 2003 and 2004 for 500 $\times$ 500 meter grids across Germany. Panel (b) displays the distribution of days across Germany over time for 500 $\times$ 500 meter grids. Source: Global Snow Pack dataset.

Figure 3.A.2.2 : Days of snow and catchment areas in Bonn

Notes: This graph displays the variation in days of snow between 2003 and 2004 for 500 × 500 meter grids in hospital catchment areas in the city of Bonn. Source: Global Snow Pack dataset.

Figure 3.A.2.3 : Population density and catchment areas in Bonn

Notes: This graph displays the variation in the population density in hospital catchment areas in the city of Bonn in 2006 based on 1,000 × 1,000 meter grids. Source: GEOSTAT.



(a) Raw correlation

(b) Correlation with municipality fixed effects

Figure 3.A.2.4 : Hospital resources and days of snow

Notes: Panel (a) shows a binscatters of the raw correlation between the number of beds and the number and days of snow in 2004. In Panel (b), I control for municipality fixed effects.

3.A.3 Additional Results

Table 3.A.3.1: Alternative IV results: The effect of treatment prices on input and output

	OLS (1)	Reduced form (2)	IV-2SLS (3)
A. Input			
Physicians	-0.002***	-0.143**	-0.004**
	(0.000)	(0.050)	(0.002)
Nurses	-0.005***	-0.185**	-0.007**
	(0.002)	(0.071)	(0.003)
Special equipment	-0.001	-0.175	-0.001
	(0.000)	(0.099)	(0.001)
B. Output			
Treatment range	0.003***	-0.126***	-0.006***
	(0.001)	(0.035)	(0.002)
Treatment volume	0.004***	0.101***	0.009**
	(0.001)	(0.031)	(0.004)
C. Structure			
Private	-0.003**	-0.197**	-0.006*
	(0.001)	(0.089)	(0.003)
Mergers	-0.002**	-0.323***	-0.005*
	(0.001)	(0.061)	(0.002)
Closed	-0.001*	-0.287***	-0.001
	(0.000)	(0.043)	(0.000)
First-stage:			
Snow instrument			44.422***
			(9.467)
F-statistic			43.737
N	8626	8626	
Controls:			
Year FE	Yes	Yes	
Municipality FE	Yes	Yes	
Hospital characteristics	Yes	Yes	
Weather characteristics	Yes	Yes	
Municipality characteristics	Yes	Yes	
Linear time trends	Yes	Yes	

Notes: This table displays the IV results based on the alternative catchment area definition. Column (1) reproduces OLS results from Column (5) in Table 3.2. Column (2) reports the reduced-form coefficients of separate regressions of the outcomes on the instrument and controls. The main IV results are displayed in Column (3). Column (3) additionally reports the first-stage coefficients of the base rate regressed on the instrument and all controls mentioned in section 3.3.1 and the corresponding F-statistics. Standard errors are displayed in parentheses and clustered at the county level. Significance level: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 3.A.3.2: Placebo test: F-statistic for different years

	2004	2005	2006
	(1)	(2)	(3)
F-statistic	47.238	4.201	2.236
N	8626	8626	8626
Controls:			
Year FE	Yes	Yes	Yes
Municipality FE	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes
Weather characteristics	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes
Linear time trends	Yes	Yes	Yes

Notes: This table displays the F-statistic of different instruments. In each Column, I construct the instrument base on the deviation of snowfall between the year listed at the top and its corresponding previous year. Column (1) reproduces the F-statistic of Table 3.3. Standard errors are displayed in parentheses and clustered at the county level. Significance level: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 3.A.3.3: Placebo test: F-statistic for different hospital samples

	General	Special
	(1)	(2)
F-statistic	47.238	12.201
N	8626	8626
Controls:		
Year FE	Yes	Yes
Municipality FE	Yes	Yes
Hospital characteristics	Yes	Yes
Weather characteristics	Yes	Yes
Municipality characteristics	Yes	Yes
Linear time trends	Yes	Yes

Notes: This table displays the F-statistic of different instruments. Column (1) reproduces the F-statistic of Table 3.3. In Column (2), I use the sample of the excluded special hospitals. Standard errors are displayed in parentheses and clustered at the county level. Significance level: $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table 3.A.3.4: Adjusted R^2 for OLS Results.

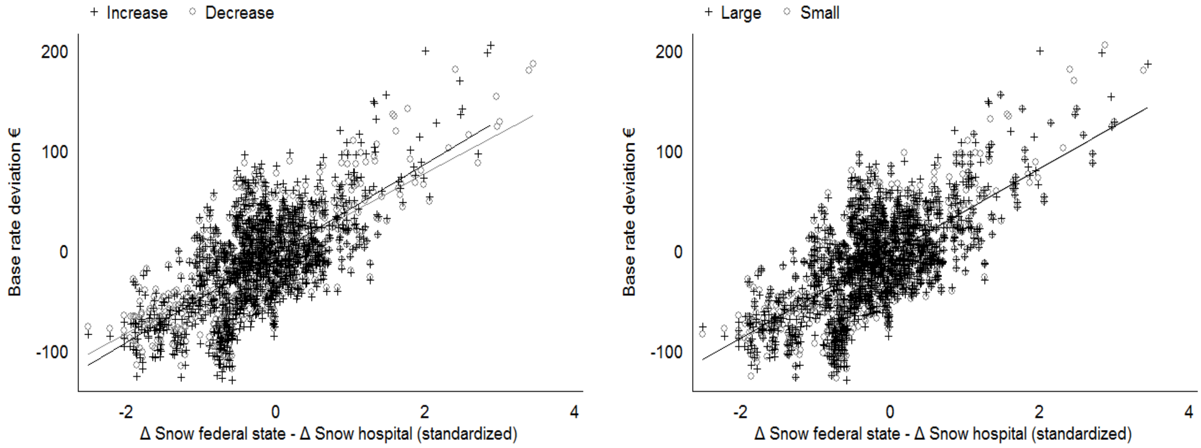
	(1)	(2)	(3)	(4)	(5)
A. Input					
Physicians	0.02	0.20	0.25	0.26	0.29
Nurses	0.01	0.18	0.21	0.21	0.24
Special equipment	0.02	0.19	0.20	0.21	0.23
B. Output					
Treatment range	0.03	0.17	0.19	0.22	0.27
Treatment volume	0.03	0.21	0.22	0.24	0.28
C. Structure					
Private	0.02	0.15	0.15	0.16	0.19
Merged	0.02	0.18	0.18	0.19	0.22
Closed	0.11	0.12	0.11	0.13	0.14
N	8626	8626	8626	8626	8626
Controls:					
Year FE	No	Yes	Yes	Yes	Yes
Municipality FE	No	Yes	Yes	Yes	Yes
Hospital characteristics	No	No	Yes	Yes	Yes
Weather characteristics	No	No	No	Yes	Yes
Municipality characteristics	No	No	No	No	Yes
Linear time trends	No	No	No	No	Yes

Notes: This table displays the adjusted R^2 for the baseline results presented in Columns (1)-(5) in Table 3.2

Table 3.A.3.5: OLS Results: The effect of treatment prices on input and output for different base rate stages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Input								
Physicians								
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price increase	0.02*							
	(0.000)							
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price decrease	0.002*							
	(0.000)							
Nurses								
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price increase		0.005**						
		(0.002)						
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price decrease		0.005**						
		(0.002)						
Special equipment								
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price increase			-0.001					
			(0.001)					
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price decrease			-0.001					
			(0.001)					
B. Output								
Treatment range								
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price increase				0.003**				
				(0.001)				
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price decrease				0.003***				
				(0.001)				
Treatment volume								
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price increase					-0.004***			
					(0.001)			
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price decrease					-0.005***			
					(0.001)			
C. Structure								
Private								
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price increase						-0.003**		
						(0.001)		
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price decrease						-0.003**		
						(0.001)		
Merged								
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price increase							-0.002**	
							(0.001)	
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price decrease							-0.002*	
							(0.001)	
Closed								
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price increase								-0.001
								(0.001)
$(br_{i_{2004}} - br_{ft_{2004}}) \times$ price decrease								-0.001
								(0.001)
N	8626	8626	8626	8626	8626	8626	8626	8626
Adj R^2	0.29	0.24	0.23	0.27	0.28	0.19	0.22	0.14
Controls:								
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linear time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table displays the estimates from OLS regressions of input and output factors on a treatment price increase that is interacted with an indicator that equals unity if a hospital base rate in 2004 was above the federal hospital base rate. Standard errors, clustered at the county level, are displayed in parentheses. Significance levels: $*** : p < 0.01, ** : p < 0.05, * : p < 0.1.$



(a) By treatment price change

(b) By hospital size

Figure 3.A.3.1 : First-stage correlation by sub-group

Notes: Notes: This figure displays the scatter plots of first-stage regressions split between hospitals with increasing and decreasing treatment prices or large and small hospitals.

Summary and Overall Conclusion

Modern health care systems are characterized by pronounced prevention and cost-optimized treatments. This dissertation offers novel empirical evidence on how useful such measures can be. The first chapter analyzes how radiation, a main pollutant in health care, can negatively affect cognitive health. The second chapter focuses on the effect of Low Emission Zones on public heath, as air quality is the major external source of health problems. Both chapters point out potentials for preventive measures. Finally, chapter three studies how changes in treatment prices affect the reallocation of hospital resources. In the following, I briefly summarize each chapter and discuss implications for health care systems as well as other policy areas.

Based on the National Educational Panel Study that is linked to data on radiation, chapter one shows that radiation can have negative long-term effects on cognitive skills, even at subclinical doses. Exploiting arguably exogenous variation in soil contamination in Germany due to the Chernobyl disaster in 1986, the findings show that people exposed to higher radiation perform significantly worse in cognitive tests 25 years later. Identification is ensured by abnormal rainfall within a critical period of ten days. The results show that the effect is stronger among older cohorts than younger cohorts, which is consistent with radiation accelerating cognitive decline as people get older. On average, a one-standard-deviation increase in the initial level of CS137 ($\approx$ 30 chest x-rays) is associated with a decrease in the cognitive skills by 4.1 percent of a standard deviation ($\approx$ 0.05 school years). Chapter one shows that sub-clinical levels of radiation can have negative consequences even after early childhood. This is of particular importance because most of the literature focuses on exposure very early in life, often during pregnancy. However, population exposed after birth is over 100 times larger. These results point to substantial external human capital costs of radiation which can be reduced by choices of medical procedures. There is a large potential for reductions because about one-third of all CT scans are assumed to be not medically justified (Brenner and Hall, 2007). If people receive unnecessary CT scans because of economic incentives, this chapter points to additional external costs of health care policies. Furthermore, the results can inform the cost-benefit trade-off for medically indicated procedures.

Chapter two provides evidence about the effectiveness of Low Emission Zones. Low Emission Zones are typically justified by improvements in population health. However, there is little evidence about the potential health benefits from policy interventions aiming at improving air quality in inner-cities. The chapter ask how the coverage of Low Emission Zones affect air pollution and hospitalization, by exploiting variation in the roll out of Low Emission Zones in Germany. It combines information on the geographic coverage of Low Emission Zones with rich panel data on the universe of German hospitals over the period from 2006 to 2016 with precise information on hospital locations and the annual frequency of detailed diagnoses.

In order to establish that our estimates of Low Emission Zones' health impacts can indeed be attributed to improvements in local air quality, we use data from Germany's official air pollution monitoring system and assign monitor locations to Low Emission Zones and test whether measures of air pollution are affected by the coverage of a Low Emission Zone. Results in chapter two confirm former results showing that the introduction of Low Emission Zones improved air quality significantly by reducing NO2 and PM10 concentrations. Furthermore, the chapter shows that hospitals which catchment areas are covered by a Low Emission Zone, diagnose significantly less air pollution related diseases, in particular by reducing the incidents of chronic diseases of the circulatory and the respiratory system. The effect is stronger before 2012, which is consistent with a general improvement in the vehicle fleet's emission standards. Depending on the disease, a one-standard-deviation increase in the coverage of a hospitals catchment area covered by a Low Emission Zone reduces the yearly number of diagnoses up to 5 percent. These findings have strong implications for policy makers. In 2015, overall costs for health care in Germany were around 340 billion euros, of which 46 billion euros for diseases of the circulatory system, making it the most expensive type of disease caused by 2.9 million cases (Statistisches Bundesamt, 2017b). Hence, reductions in the incidence of diseases of the circulatory system may directly reduce society's health care costs.

Whereas chapter one and two study the demand-side in health care markets and thus preventive potential, chapter three analyzes the supply-side. By exploiting the same hospital panel data set as in chapter two, chapter three studies the effect of treatment price shocks on the reallocation of hospital resources in Germany. Starting in 2005, the implementation of the German-DRG-System led to general idiosyncratic treatment price shocks for individual hospitals. Thus far there is little evidence of the impact of general price shocks on the reallocation of hospital resources. Additionally, I add to the exiting literature by showing that price shocks can have persistent effects on hospital resources even when these shocks vanish. However, simple OLS regressions would underestimate the true effect, due to endogenous treatment price shocks. I implement a novel instrument variable strategy that exploits the exogenous variation in the number of days of snow in hospital catchment areas. A peculiarity of the reform allowed variation in days of snow to have a persistent impact on treatment prices.

I find that treatment price increases lead to increases in input factors such as nursing staff, physicians and the range of treatments offered but to decreases in the treatment volume. This indicates supplier-induced demand. Furthermore, the probability of hospital mergers and privatization decreases. Structural differences in pre-treatment characteristics between hospitals enhance these effects. For instance, private and larger hospitals are more affected. IV estimates reveal that OLS results are biased towards zero in almost all dimensions because structural hospital differences are correlated with the reallocation of hospital resources.

These results are important for several reasons. The G-DRG-Reform led to a persistent polarization of hospital resources, as some hospitals were exposed to treatment price increases, while others experienced reductions. If hospitals increase the treatment volume as a response to price reductions by offering unnecessary therapies, it has a negative impact on population well-being and public spending. However, results show a decrease in the range of treatments if prices decrease. Hospitals might specialize more, thus attracting more patients. From a policy perspective it is important to evaluate if such changes in the range of treatments jeopardize an adequate nationwide provision of treatments. Furthermore, the results show a decrease in the number of nurses and physicians if prices decrease. This could partly explain the nursing crisis in German hospitals. However, since hospitals specialize more they might be able to realize efficiency gains which justify reductions in input factors without loses in quality. Further research is necessary to provide evidence for the impact of the G-DRG-Reform on health care quality. Another important aspect are changes in the organizational structure. Many public hospitals have been privatized or merged. The findings show that this is at least partly driven by the G-DRG-Reform. This can again lead to a lack in services offered in some regions if merged hospitals specialize more or if hospitals are taken over by ecclesiastical organizations which do not provide all treatments due to moral conviction.

Overall, this dissertation reveals large potential for preventive health care measures and helps to explain reallocation processes in the hospital sector if treatment prices change. Furthermore, its findings have potentially relevant implications for other areas of public policy. Chapter one identifies an effect of low dose radiation on cognitive health. As mankind is searching for new energy sources, nuclear power is becoming popular again. However, results of chapter one point to substantial costs of nuclear energy which have not been accounted yet. Chapter two finds strong evidence that air quality improvements by Low Emission Zones translate into health improvements, even at relatively low levels of air pollution. These findings may, for instance, be of relevance to design further policies targeted at air pollution such as diesel bans. As pointed out in chapter three, the implementation of DRG-Systems may have unintended side-effects on the reallocation of hospital resources. This may also apply to other providers in the health care sector such as resident doctors.

German Summary

Moderne Gesundheitssysteme zeichnen sich sowohl durch eine ausgeprägte Prävention als auch durch kostenoptimierte Behandlungen aus. Diese Dissertation bietet neue empirische Erkenntnisse darüber, wie nützlich solche Maßnahmen sein können. Das erste Kapitel analysiert, wie Strahlung, ein Hauptschadstoff im Gesundheitswesen, die kognitive Gesundheit negativ beeinflussen kann. Das zweite Kapitel konzentriert sich auf die Auswirkungen von Umweltzonen auf die öffentliche Gesundheit, da die Luftqualität die wichtigste externe Quelle für Gesundheitsprobleme ist. Beide Kapitel zeigen Potenziale für präventive Maßnahmen auf. Schließlich wird in Kapitel drei untersucht, wie sich Änderungen von Behandlungspreisen auf die Reallokation von Krankenhausressourcen auswirken. Im Folgenden fasse ich jedes Kapitel kurz zusammen und diskutiere die Relevanz für Gesundheitssysteme und andere Politikbereiche.

Basierend auf dem Nationalen Bildungspanel, welches wir mit Strahlungsdaten verknüpfen, zeigt Kapitel eins, dass Strahlung, auch in geringen Dosen, negative Langzeitwirkungen auf die kognitiven Fähigkeiten haben kann. Dazu nutzen wir die exogene Variation der Bodenkontamination in Deutschland nach der Tschernobyl-Katastrophe von 1986. Die Ergebnisse zeigen, dass Menschen, die einer höheren Strahlung ausgesetzt waren, 25 Jahre später in kognitiven Tests deutlich schlechter abschneiden. Die Identifizierung wird durch anormale Niederschläge innerhalb eines kritischen Zeitraums von zehn Tagen nach dem Reaktorunfall gewährleistet. Die Ergebnisse zeigen, dass der Effekt bei älteren Kohorten stärker ist als bei jüngeren Kohorten, was mit der Theorie übereinstimmt, dass Strahlung den altersbedingten Rückgang der kognitiven Leistungsfähigkeit beschleunigt. Im Durchschnitt ist eine Erhöhung des Anfangsniveaus von CS137 um eine Standardabweichung ($\approx$ 30 Thoraxröntgenaufnahmen) mit einer Abnahme der kognitiven Fähigkeiten um 4,1 Prozent einer Standardabweichung verbunden ($\approx$ 0.05 Schuljahre).

Die Ergebnisse in Kapitel eins zeigen, dass geringe Strahlungswerte auch nach der frühen Kindheit negative Folgen haben können. Dies ist von besonderer Bedeutung, da sich der Großteil der Literatur auf die Exposition in sehr frühen Lebensphasen konzentriert, oft während der Schwangerschaft. Die nach der Geburt exponierte Bevölkerung ist jedoch über 100-mal größer. Diese Ergebnisse deuten auf erhebliche externe Humankapitalkosten der Strahlung hin, die zum Beispiel durch die Wahl medizinischer Verfahren reduziert werden können. Es bestehen große Reduktionspotenziale, da beispielsweise etwa ein Drittel aller CT-Scans als medizinisch nicht gerechtfertigt angesehen werden (Brenner and Hall, 2007). Unter der Annahme, dass Menschen aufgrund wirtschaftlicher Anreize unnötige CT-Scans erhalten, weist dieses Kapitel auf zusätzliche externe Kosten von Gesundheitsmaßnahmen für die Gesundheit der Patienten hin. Außerdem erweitern die Ergebnisse die Informationsgrundlage für Risiko-Nutzen-Abwägungen medizinischer Behandlungen.

Kapitel zwei liefert Belege für die Wirksamkeit von Umweltzonen. Umweltzonen sind in der Regel durch eine Verbesserung der Gesundheit der Bevölkerung gerechtfertigt. Es gibt jedoch wenig Belege für den gesundheitlichen Nutzen solcher politischen Maßnahmen zur Verbesserung der Luftqualität in Innenstädten. In dem Kapitel analysieren wir, wie sich die Ausdehnung der Umweltzonen auf die Luftverschmutzung und die Krankenhausaufenthalte auswirkt. Dazu nutzen wir zeitliche Unterschiede bei der Einführung der Umweltzonen in Deutschland. Hierfür kombinieren wir Informationen über die geografische Abdeckung der Umweltzonen mit umfangreichen Paneldaten von allen deutschen Krankenhäusern im Zeitraum von 2006 bis 2016. Die Krankenhausdaten enthalten präzise Informationen über Krankenhausstandorte und die jährliche Häufigkeit detaillierter Diagnosen.

Um sicherzustellen, dass unsere Schätzungen der gesundheitlichen Auswirkungen der Umweltzonen auf eine Verbesserung der lokalen Luftqualität zurückzuführen sind, verwenden wir Daten aus dem offiziellen deutschen Luftmessnetz und ordnen den Umweltzonen Monitorstandorte zu. Hierdurch prüfen wir, ob Messungen der Luftverschmutzung durch die Abdeckung einer Umweltzone beeinflusst werden. Die Ergebnisse in Kapitel zwei bestätigen frühere Ergebnisse, die zeigen, dass die Einführung von Umweltzonen die Luftqualität durch die Reduzierung der NO2- und PM10-Konzentrationen deutlich verbessert. Darüber hinaus zeigt das Kapitel, dass Krankenhäuser, deren Einzugsgebiete in eine Umweltzone fallen, deutlich weniger durch Luftverschmutzung bedingte Krankheiten diagnostizieren, insbesondere durch die Verringerung von chronischen Krankheiten des Kreislaufs und der Atemwege. Der Effekt ist vor 2012 stärker, was sich durch eine allgemeine Verbesserung der Abgasnormen für Fahrzeuge erklären lässt. Wird der Anteil eines Krankenhauseinzugsgebiets mit Umweltzone um eine Standardabweichung erhöht, reduziert sich die jährliche Anzahl der entsprechenden Diagnosen um bis zu 5 Prozent.

Diese Ergebnisse sind für politische Entscheidungsträger von Bedeutung. Im Jahr 2015 gab Deutschland 46 Milliarden Euro für Herz-Kreislauf-Erkrankungen aus, die teuerste Krankheitsform mit 2,9 Millionen Fällen. Die Reduktion von Herz-Kreislauf-Erkrankungen würde die Gesundheitskosten der Gesellschaft unmittelbar senken.

Während Kapitel eins und zwei die Nachfrageseite in Gesundheitsmärkten und damit das Präventionspotenzial untersuchen, analysiert Kapitel drei die Angebotsseite. Unter Verwendung desselben Krankenhauspaneldatensatzes wie in Kapitel zwei untersuche ich die Auswirkungen von veränderten Behandlungspreisen auf Krankenhausressourcen in Deutschland. Ab 2005 wurden die Behandlungspreise für Patienten durch das G-DRG-

System umgewandelt, was bis 2010 zu allgemeinen idiosynkratischen Preisschocks für einzelne Krankenhäuser führte. Die vorhandene Literatur bildet den Effekt von allgemeinen Preisschocks auf Krankenhausressourcen jedoch nur unzureichend ab. Außerdem erweitere ich die vorhandene Literatur indem ich zeige, dass Preisänderungen auch dann langfristige Auswirkungen auf Krankenhausressourcen haben können, wenn die Preisschocks verschwinden. Einfache OLS-Regressionen würden den wahren Effekt aufgrund von endogenen Preisschocks unterschätzen. Unter Zuhilfenahme von hochauflösenden Satellitendaten nutze ich eine Instrumentenvariablenstrategie, welche exogene Schwankungen der Wetterbedingungen im Einzugsbereich von Krankenhäusern nutzt. Eine Besonderheit der Reform führt dazu, dass Abweichungen der Wetterbedingungen zum Zeitpunkt der Reformeinführung einen nachhaltigen Einfluss auf die Behandlungspreise hatten.

Die Ergebnisse zeigen, dass Preiserhöhungen im Laufe der Zeit zu einem Anstieg des Pflegepersonals, von Ärzten und der Vielfalt der angebotenen Behandlungen führen, aber zu einem Rückgang des Behandlungsvolumens. Bei privaten und größeren Krankenhäusern sind die Auswirkungen stärker. Darüber hinaus sinkt die Wahrscheinlichkeit von Krankenhausfusionen und Privatisierungen. IV-Ergebnisse zeigen gegen Null verzerrte OLS-Schätzungen in fast allen Dimensionen, da strukturelle Krankenhausunterschiede mit der Reallokation von Ressourcen korreliert sind.

Diese Ergebnisse sind aus mehreren Gründen wichtig. Die G-DRG-Reform führte zu einer anhaltenden Polarisation von Krankenhausressourcen, da Krankenhäuser sowohl Preisanstiege als auch Preissenkungen erfuhren. Wenn Krankenhäuser das Behandlungsvolumen durch unnötige Therapien erhöhen, hat das negative Auswirkungen auf die öffentliche Gesundheit der Bevölkerung und die öffentlichen Ausgaben. Andererseits zeigen die Ergebnisse einen Rückgang der Bandbreite der angebotenen Behandlungen bei sinkenden Preisen. Krankenhäuser könnten sich stärker spezialisieren und so mehr Patienten anziehen. Aus politischer Sicht ist es wichtig zu beurteilen, ob solche Veränderungen in der Vielfalt der angebotenen Behandlungen eine angemessene flächendeckende Versorgung gefährden. Des Weiteren zeigen die Ergebnisse einen Rückgang der Zahl der Krankenschwestern und Ärzte, wenn die Preise sinken. Dies könnte die Pflegekrise, welche die Bundesregierung in Deutschland beschreibt, teilweise erklären. Da sich die Krankenhäuser jedoch stärker spezialisieren, können sie möglicherweise Effizienzsteigerungen erzielen, die eine Verringerung der Inputfaktoren rechtfertigen, ohne an Qualität zu verlieren. Weitere Untersuchungen sind notwendig, um die Auswirkungen auf die Qualität der Gesundheitsversorgung nachzuweisen. Ein weiterer wichtiger Aspekt sind Veränderungen in der Organisationsstruktur. Viele öffentliche Krankenhäuser werden privatisiert oder mit anderen Krankenhäusern fusioniert. Meine Ergebnisse zeigen, dass dies zumindest teilweise auf die G-DRG-Reform zurückzuführen ist. Dies kann zu einem Mangel an angebotenen Behandlungen in einigen Regionen führen, wenn sich fusionierte Krankenhäuser spezialisieren oder wenn Krankenhäuser von kirchlichen Organisationen übernommen werden, die aus moralischen Gründen nicht alle Behandlungen anbieten.

Insgesamt unterstreicht diese Dissertation das große Potenzial von Gesundheitsvorsorgemaßnahmen und hilft, Reallokationsprozesse im Krankenhaussektor zu erklären. Darüber hinaus haben die Ergebnisse potenziell relevante Auswirkungen auf andere Bereiche der Politik. Kapitel Eins identifiziert einen Einfluss von geringer Radioaktivität auf die kognitive Gesundheit. Auf der Suche nach neuen Energiequellen wird die Kernenergie wieder populär. Die Ergebnisse von Kapitel Eins deuten jedoch auf erhebliche Kosten von Kernenergie hin, die in der aktuellen Debatte noch nicht berücksichtigt wurden. Kapitel Zwei findet starke Hinweise darauf, dass die Verbesserung der Luftqualität durch Umweltzonen, selbst bei relativ geringer Luftverschmutzung, zu einer Verbesserung der Gesundheit führt. Diese Ergebnisse können für die Einführung weiterer Maßnahmen zur Bekämpfung der Luftverschmutzung von Bedeutung sein, wie beispielsweise Fahrverbote für Dieselfahrzeuge. Wie in Kapitel Drei dargelegt, kann die Einführung von DRG-Systemen unbeabsichtigte Effekte bei der Reallokation von Krankenhausressourcen haben. Dies kann auch für andere Anbieter im Gesundheitswesen wie niedergelassene Ärzte gelten.

Bibliography

Acemoglu, D. and Finkelstein, A. (2008). Input and technology choices in regulated industries: Evidence from the health care sector. *Journal of Political Economy* 116: 837–880.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J. and Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Jama* 288: 1987–1993.

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

Almond, D. and Currie, J. (2013). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives* 25(3): 153–172.

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

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103: 1481–1495.

Antoni, M., Drasch, K., Kleinert, C., Matthes, B., Ruland, M. and Trahms, A. (2011). Working and learning in a changing world: Part I: Overview of the study-March 2011 (Second, updated version) .

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

Arrow, K. J. (1978). Uncertainty and the welfare economics of medical care. In *Uncertainty in Economics*. Elsevier, 345–375.

Auger, N., Potter, B. J., Smargiassi, A., Bilodeau-Bertrand, M., Paris, C. and Kosatsky, T. (2017). Association between quantity and duration of snowfall and risk of myocardial infarction. *CMAJ* 189: E235–E242.

Augurzky, B., Gülker, R., Krolop, S., Schmidt, C. M., Schmidt, H., Schmitz, H. and Terkatz, S. (2010). *Krankenhaus Rating Report 2010: Licht und Schatten*. 59. RWI Materialien.

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

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

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B* 57: 289–300.

Berendes, K., Weinert, S., Zimmermann, S. and Artelt, C. (2013). Assessing language indicators across the lifespan within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online* 5: 15.

Björnstig, U., Björnstig, J. and Dahlgren, A. (1997). Slipping on ice and snow–elderly women and young men are typical victims. *Accident Analysis & Prevention* 29: 211–215.

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* .

Block, M. L. and Calderon-Garciduenas, L. (2009). Air pollution: mechanisms of neuroinflammation and CNS disease. *Trends in Neurosciences* 32(9).

Bloom, D. E., Canning, D. and Sevilla, J. (2001). The effect of health on economic growth: theory and evidence .

Blossfeld, H.-P., Rossbach, H.-G. and Maurice, J. von (2011). Education as a Life-long Process - The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft* 14.

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

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

Brunner, M., Lang, F. R. and Lüdtke, O. (2014). Erfassung der fluiden kognitiven Leistungsfähigkeit über die Lebensspanne im Rahmen der National Educational Panel Study: Expertise . *NEPS Working Paper* .

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

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

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

Busse, R. and Quentin, W. (2011). *Moving towards transparency, efficiency and quality in hospitals: conclusions and recommendations*. New York: Open University Press.

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

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

Chalkley, M. and Malcomson, J. M. (2000). Government purchasing of health services. In *Handbook of health economics*. Elsevier, *1*, 847–890.

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

Clemens, J. and Gottlieb, J. D. (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review* 104: 1320–49.

Cohen, J. T., Neumann, P. J. and Weinstein, M. C. (2008). Does preventive care save money? health economics and the presidential candidates. *New England Journal of Medicine* 358: 661–663.

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.

Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review* 95: 1525–1547.

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.

Deschênes, O., Greenstone, M. and Shapiro, J. S. (2017). Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program. *American Economic Review* 107: 2958–2589.

Dietz, A. J., Kuenzer, C. and Dech, S. (2015). Global snowpack: A new set of snow cover parameters for studying status and dynamics of the planetary snow cover extent. *Remote sensing letters* 6: 844–853.

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.

Eisenberg, D. and Warner, K. E. (2005). Effects of snowfalls on motor vehicle collisions, injuries, and fatalities. *American journal of public health* 95: 120–124.

Environmental Protection Agency (2016). Integrated Science Assessment for Oxides of Nitrogen-Health Criteria.

Epstein, D. and Mason, A. (2006). Costs and prices for inpatient care in england: mirror twins or distant cousins? *Health Care Management Science* 9: 233–242.

Ethikrat (2016). Patientenwohl als ethischer Maßstab für das Krankenhaus .

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

European Commission (2018). Air quality: Commission takes action to protect citizens from air pollution .

European Environmental Agency (2018). Air Quality in Europe - 2018 Report. *EEA Report* No. 12/2018.

Evans, R. G. (1974). Supplier-induced demand: some empirical evidence and implications. In *The economics of health and medical care*. Springer, 162–173.

Farsi, M. and Filippini, M. (2008). Effects of ownership, subsidization and teaching activities on hospital costs in switzerland. *Health economics* 17: 335–350.

Feldstein, M. S. (1973). The welfare loss of excess health insurance. *Journal of Political Economy* 81: 251–280.

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

Filistrucchi, L. and Prüfer, J. (2018). Faithful strategies: how religion shapes nonprofit management. *Management science* 65: 188–208.

Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of medicare. *The quarterly journal of economics* 122: 1–37.

Franklin, B. A., Hogan, P., Bonzheim, K., Bakalyar, D., Terrien, E., Gordon, S. and Timmis, G. C. (1995). Cardiac demands of heavy snow shoveling. *Jama* 273: 880–882.

Friedrich, J. and Beivers, A. (2008). Patientenwege ins Krankenhaus: Räumliche Mobilität bei Elektiv-und Notfallleistungen am Beispiel der Hüftendoprothesen. *Krankenhausreport* 9: 155–180.

Fürstenberg, T., Laschat, M., Zich, K., Klein, S., Gierling, P., Nolting, H. and Schmidt, T. (2013). G-DRG-Begleitforschung gemäß § 17b Abs. 8 KHG. Endbericht des dritten Forschungszyklus (2008 bis 2010). *Untersuchung im Auftrag des deutschen DRG-Instituts (InEK). Düsseldorf: Deutsche Krankenhaus Verlagsgesellschaft* .

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

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

Gehrsitz, M. (2017). The effect of low emission zones on air pollution and infant health. *Journal of Environmental Economics and Management* 83: 121–144.

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

Gowrisankaran, G., Nevo, A. and Town, R. (2015). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review* 105: 172–203.

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.

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

Graff Zivin, J. and Neidell, M. (2018). Air pollution's hidden impacts. *Science* 359: 39–40.

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

Guterman, S. and Dobson, A. (1986). Impact of the medicare prospective payment system for hospitals. *Health Care Financing Review* 7: 97.

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

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

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

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

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

Herr, A., Schmitz, H. and Augurzky, B. (2011). Profit efficiency and ownership of German hospitals. *Health Economics* 20: 660–674.

Hodgkin, D. and McGuire, T. G. (1994). Payment levels and hospital response to prospective payment. *Journal of Health Economics* 13: 1–29.

Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B. and D., K. J. (2013). Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environmental Health* 12:43.

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

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

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

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

Janke, K. (2014). Air pollution, avoidance behaviour and children's respiratory health: Evidence from England. *Journal of Health Economics* 38: 23 – 42.

Jiang, W., Boltze, M., Groer, S. and Scheuvens, D. (2017). Impacts of low emission zones in Germany on air pollution levels. *Transportation Research Procedia* 25: 3370 – 3382.

Kampa, M. and Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution* 151: 362–367.

Karagulian, F., Belis, C. A., Dora, C. F. C., Prüss-Ustün, A. M., Bonjour, S., Adair-Rohani, H. and Amann, M. (2015). Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmospheric Environment* 120: 475 – 483.

Karlsson, M. and Ziebarth, N. R. (2018). Population health effects and health-related costs of extreme temperatures: Comprehensive evidence from Germany. *Journal of Environmental Economics and Management* 91: 93 – 117.

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

Klauber, J. (2006). Krankenhaus-Report 2006: Schwerpunkt: Krankenhausmarkt im Umbruch .

Klauber, J., Geraedts, M., Friedrich, J. and Wasem, J. (2011). Krankenhaus-Report 2011. Schwerpunkt: Qualität durch Wettbewerb.

Klauber, J., Geraedts, M., Friedrich, J. and Wasem, J. (2018). *Krankenhaus-Report 2018: Schwerpunkt: Bedarf und Bedarfsgerechtigkeit-Mit Online-Zugang*. Georg Thieme Verlag.

Klauber, J., Robra, B. P. and Schellschmid, H. (2008). Krankenhausreport 2008: Schwerpunkt: Versorgungszentren .

Klauber, J., Wasem, J., Friedrich, J. and Geraedts, M. (2015). Krankenhaus-Report 2015: Schwerpunkt: Strukturwandel .

Kraftfahrt-Bundesamt (2018). Fahrzeugzulassungen (FZ) Bestand an Kraftfahrzeugen nach Umwelt-Merkmalen .

Krishnan, R. (2001). Market restructuring and pricing in the hospital industry. *Journal of Health Economics* 20: 213–237.

Kristensen, T., Bogetoft, P. and Pedersen, K. M. (2010). Potential gains from hospital mergers in Denmark. *Health care management science* 13: 334–345.

Kumar, A. and Schoenstein, M. (2013). Managing Hospital Volumes: Germany and Experiences from OECD Countries. *OECD Health Working Papers* 64.

Landrigan, P. J., Fuller, R., Acosta, N. J., Adeyi, O., Arnold, R., Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., Breysse, P. N. et al. (2017). The Lancet Commission on pollution and health. *The Lancet* .

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.

Levin, D. C., Rao, V. M., Parker, L., Frangos, A. J. and Sunshine, J. H. (2008). Ownership or leasing of MRI facilities by nonradiologist physicians is a rapidly growing trend. *Journal of the American College of Radiology* 5: 105–109.

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

Maccini, S. and Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review* 99: 1006–26.

Maciosek, M. V., Coffield, A. B., Edwards, N. M., Flottemesch, T. J., Goodman, M. J. and Solberg, L. I. (2006). Priorities among effective clinical preventive services: results of a systematic review and analysis. *American journal of preventive medicine* 31: 52–61.

Malina, C. and Scheffler, F. (2015). The impact of low emission zones on particulate matter concentration and public health. *Transportation Research Part A: Policy and Practice* 77: 372–385.

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

McGuire, T. G. (2000). Physician agency. In *Handbook of health economics*. Elsevier, *1*, 461–536.

McLafferty, S. L. (2003). GIS and Health Care. *Annual Review of Public Health* 24: 25–42.

Mokdad, A. H., Marks, J. S., Stroup, D. F. and Gerberding, J. L. (2004). Actual causes of death in the united states, 2000. *Jama* 291: 1238–1245.

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.

Morfeld, P., Groneberg, D. A. and Spallek, M. F. (2014). Effectiveness of Low Emission Zones: Large Scale Analysis of Changes in Environmental NO2, NO and NOx Concentrations in 17 German Cities. *PLOS ONE* 9(8).

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

NCRP (2009). Ionizing Radiation Exposure of the Population of the United States 160.

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M. and Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine* 346: 1715–1722.

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

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

OECD (2016). Radiological Protection Science and Application .

OECD (2018). Spending on health: Latest trends .

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

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

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

Peltzman, S. (2000). Prices rise faster than they fall. *Journal of political economy* 108: 466–502.

Peters, H. P., Albrecht, G., Hennen, L. and Stegelmann, U. (1987). Die Reaktionen der Bevölkerung auf die Ereignisse in Tschernobyl - Ergebnisse einer Befragung. *Forschungszentrum Jülich GmbH Zentralbibliothek* .

Pope III, C. A. (2000). Epidemiology of Fine Particulate Air Pollution and Human Health: Biologic Mechanisms and who is at Risk? *Environmental Health Perspectives* 108: 713–723.

Pope III, C. A. and Dockery, D. W. (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect. *Journal of the Air and Waste Management Association* 56: 709–742.

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.

Ralis, Z. (1981). Epidemic of fractures during period of snow and ice. *British Medical Journal* 282: 603–605.

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

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

Rice, T. H. (1983). The impact of changing medicare reimbursement rates on physician-induced demand. *Medical care* : 803–815.

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.

Salm, M. and Wübker, A. (2015). Do hospitals respond to increasing prices by supplying fewer services? *Available at SSRN 2640116* .

Salm, M. and Wübker, A. (2018). *Do higher hospital reimbursement prices improve quality of care?*. 779. Ruhr Economic Papers.

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

Saltman, R. B. and Figueras, J. (1998). Analyzing the evidence on european health care reforms: Experience in western european health care systems suggests lessons for reform in the united states, according to a major international comparison by the world health organization. *Health affairs* 17: 85–108.

Salvo, A., He, J. and Gouveia, N. (2018). External Effects of Diesel Trucks Circulating Inside the Sao Paulo Megacity. *Journal of the European Economic Association* forthcoming.

Sánchez-Martínez, F., Abellán-Perpiñán, J.-M., Martínez-Pérez, J.-E. and Puig-Junoy, J. (2006). Cost accounting and public reimbursement schemes in spanish hospitals. *Health Care Management Science* 9: 225–232.

Schlenker, W. and Walker, W. R. (2016). Airports, Air Pollution, and Contemporaneous Health. *The Review of Economic Studies* 83: 768–809.

Schneider, A., Cyrys, J., Breitner, S., Kraus, U., Peters, A., Diegmann, V. and Neunhaeuserer, L. (2018). Quantifizierung von umweltbedingten Krankheitslasten aufgrund der Stickstoffdioxid Exposition in Deutschland. *Im Auftrag des Umweltbundesamtes* .

Schnittjer, I. and Duchhardt, C. (2015). Mathemathcal Competence:Framework and Exemplary Test Items .

Schreyögg, J., Bäuml, M., Krämer, J., Dette, T., Busse, R. and Geissler, A. (2014). Forschungsauftrag zur mengenentwicklung nach § 17b abs. 9 khg. *Hamburg Center for Health Economics (hche), Hamburg* .

Schreyögg, J., Tiemann, O. and Busse, R. (2006). Cost accounting to determine prices: how well do prices reflect costs in the German DRG-system? *Health Care Management Science* 9: 269–279.

Schuurman, N., Fiedler, R. S., Grzybowski, S. C. W. and Grund, D. (2006). Defining rational hospital catchments for non-urban areas based on travel-time. *International Journal of Health Geographics* .

Selbmann, H.-K. (2004). Der Qualitätsbericht ab 2005–wozu und für wen. *Das Krankenhaus* 9: 712–716.

Simeonova, E., Currie, J., Nilsson, P. and Walker, R. (2018). Congestion Pricing, Air Pollution and Children's Health. Working Paper 24410, National Bureau of Economic Research.

Sloan, F. A. and Hsieh, C.-R. (2017). *Health economics*.

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

Statistisches Bundesamt (2017a). Fallpauschalenbezogene Krankenhausstatistik (DRG-Statistik) Diagnosen, Prozeduren, Fallpauschalen und Case Mix der vollstationären Patientinnen und Patienten in Krankenhäusern. *Fachserie 12 Reihe 6.4* .

Statistisches Bundesamt (2017b). Herz-Kreislauf-Erkrankungen verursachen die höchsten Kosten. *Pressemitteilung vom 29. September 2017 - 347/17* .

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.

Tiwary, A. and Williams, I. (2018). *Air pollution: measurement, modelling and mitigation*. CRC Press.

Unruh, L. (2003). Licensed nurse staffing and adverse events in hospitals. *Medical care* : 142–152.

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

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

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

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

Vitousek, P., Aber, J., Howarth, R., Likens, G., Matson, P., Schindler, D., Schlesinger, W. and Tilman, D. (1997). Human alteration of the global nitrogen cycle: Sources and consequences. *Ecological Appplications* 7: 737–750.

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 .

WHO (2006). Air Quality Guidelines – Global Update 2005; Particulate Matter, Ozone, Nitrogen dioxide ans Sulfur dioxide. *WHO Regional Office for Europe, Copenhagen, Denmark* .

WHO (2018a). Burden of disease from the joint effects of household and ambient Air pollution for 2016. *Public Health, Social and Environmental Determinants of Health Department* .

WHO (2018b). Public spending on health: a closer look at global trends .

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

Winkelmann, I., Haubelt, R., Neumann, P. and Fields, D. (1989). Radionuclide deposition and exposure in the Federal Republic of Germany after the Chernobyl accident .

Wissenschaftliche Dienste (2014). Krankenhäuser in privater Trägerschaft - Rechtsgrundlagen, verfassungsrechtliche Vorgaben und Finanzierung. *Deutscher Bundestag* WD 9 - 3000 - 095/13.

Wolff, H. (2014). Keep Your Clunker in the Suburb: Low-emission Zones and Adoption of Green Vehicles. *The Economic Journal* 124: 481–512.

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.

Yip, W. C. (1998). Physician response to medicare fee reductions: changes in the volume of coronary artery bypass graft (cabg) surgeries in the medicare and private sectors. *Journal of health economics* 17: 675–699.

Zimmermann, S., Artelt, C. and Weinert, S. (2014). The Assessment of Reading Speed in Adults and First-Year Students .