*** First draft. Please do not cite or circulate without the permission of the author. ***

Age-specific Effects of Early Daycare on Children's Health*

Mara Barschkett

DIW Berlin

Freie Universität Berlin

Berlin School of Economics

February 14, 2022

Abstract

Over the past decades, the share of very young children in daycare has increased significantly in many OECD countries including Germany. Despite the relevance of child health for the development of (non-)cognitive skills, the effects of daycare attendance on health have received little attention in the economic literature. In this study, I investigate the impact of a large daycare expansion in Germany on age-specific mental and physical health outcomes of children. Based on administrative health records covering 90% of the German population, I employ difference-and-differences as well as event study approaches taking advantage of temporal and spatial variation in the expansion speed of daycare slots. My results provide evidence for a substitution of illness spells from the first years of elementary school to the first years of daycare. Namely, I find that early daycare attendance increases the prevalence of respiratory and infectious diseases as well as healthcare consumption when entering daycare (1–2 years), but decreases the prevalence at elementary school age. I do not find evidence for an effect of the daycare expansion on mental disorders, obesity and healthcare costs.

Keywords: child care, daycare expansion, physical health, mental health, education, health registers, difference-in-differences, event study

JEL classification: I10, I12, J13, C23

Supervised by: Prof. C. Katharina Spiess, Prof. Peter Haan

*I am grateful to the National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung, KBV) for data access and for their excellent support. The author declares that there is no conflict of interest.

mbarschkett@diw.de, DIW Berlin, Mohrenstrasse 58, 10117 Berlin

1 Introduction

Over the last two decades, the share of very young children (0–2 years) in daycare has increased significantly in many OECD countries.[1] Germany experienced one of the largest increases among all OECD countries from 17% in 2005 to 37% in 2018 (e.g. OECD, 2020). Along with this development, the body of literature studying the effects of (early) daycare attendance of children on their (non-)cognitive outcomes has grown. Previous research has shown that health is one of the most important determinants of both short- and long-term (non-)cognitive outcomes (e.g. Heckman, 2007). However, despite their relevance, the effects of (early) daycare attendance on health have received little attention in the literature.

In this paper I fill this research gap and analyze age-specific effects of early daycare[2] attendance of children on their mental and physical health. To overcome the endogeneity of the decision to attend daycare at early ages, I exploit a large-scale daycare expansion in Germany. This expansion was induced by a federal reform that introduced a legal entitlement for a daycare slot for all children aged one year and older. Following the announcement of the reform in 2007, daycare coverage of under three year old children increased from about 11.7% in 2008 to 28.8% in 2018 in West-Germany (Destatis, 2019).[3] The reform generated large temporal and spatial variations in the expansion speed of daycare slots at the county-level. Using this variation, I employ difference-in-differences and event-study approaches similar to Havnes and Mogstad (2011) and Müller and Wrohlich (2020) to identify causal effects.

The analyses are based on administrative health records covering all individuals insured through the public health system in Germany (about 90% of population) between 2009 and 2019. My sample includes children from birth cohorts 1999 to 2015 which amounts to about 11 million children. The data covers the outpatient register that contains all ambulatory care contacts including all contacts with physicians, pediatricians, and therapists. Comprehensive diagnoses by practitioners based on the International Classification of Diseases (ICD-10) are recorded for each visit. Furthermore, treatment cases and healthcare costs aggregated on quarterly level are documented.

[1] On average, the enrollment rate of 0–2 year old children in daycare increased from 21% in 2005 to 32% in 2018.

[2] The term daycare describes all forms of formal child care provided by professionals outside the family. Daycare centers in Germany are publicly funded by the state (see for instance Spiess (2008) for a description).

[3] Traditionally, labor market participation of mothers and daycare attendance of young children is substantially higher in East-Germany than in West-Germany. Therefore, the situation in East and West German counties is not comparable. For this reason, I focus on West-Germany in this project.

My study contributes to the literature in several ways. First, I estimate age-specific effects by assessing instantaneous effects on child health (age one to two) as well as short-term effects (age three to five) and longer term effects at elementary school age (age six to ten). Physical and mental health as well as non-cognitive skills during childhood are important predictors of later educational achievements, health outcomes and labor market success during adulthood (see e.g. Carneiro et al., 2007; Currie, 2020; Currie and Stabile, 2006; Heckman, 2007; Heckman et al., 2013; Peet et al., 2015). Second, my study provides novel evidence on health effects of a large-scale expansion of publicly funded, high-quality, daycare in Germany. Most previous studies have either studied large-scale but rather low-quality daycare expansions (e.g. the expansion in Quebec (e.g. Baker et al., 2008, 2019; Kottelenberg and Lehrer, 2013)) or small programs targeted on children from disadvantaged families (e.g. Perry Preschool Program (e.g. Conti et al., 2016; Heckman et al., 2010)).

Third, my detailed diagnosis data enables me to gain a comprehensive understanding of potential heterogeneity of the health effects. Most previous studies rely on survey data that contain rather broad and subjective health measures. Survey data allow to assess health and behavioral outcomes that can't be measured otherwise and usually provide a large range of socio-economic characteristics. However, to measure health and behavioral disorders, survey data are less detailed than administrative health records and are potentially subject to a reporting bias. In order to get a comprehensive understanding of the daycare effects on health, it is important to study various dimensions as different diseases could be differently affected. To the best of my knowledge, so far van den Berg and Siflinger (2020) is the only study that assesses the health effects of daycare using detailed administrative health records. Fourth, I study parental care as the counterfactual (i.e. alternative) care option[4], which differs from the setting in other countries. For example, van den Berg and Siflinger (2020) study a switch from informal to formal daycare arrangements following a daycare reform in Sweden, which represents a less drastic change than a move from home care to daycare. Lastly, due to the panel structure of the data, I observe children for 11 years and am thereby able to assess both short and longer term effects.

The few studies that have assessed the effects of (early) daycare exposure on child health are inconclusive in terms of the direction and magnitude of the effects. On the one hand,

[4]The daycare expansion was already shown to increase labor market participation of mothers with young children (Müller and Wrohlich, 2020). Thus, I compare health outcomes of children in daycare compared to children in parental care.

several authors find negative effects on physical and mental child health of the introduction of public childcare in Quebec both in the short- and in the long-run (e.g. Baker et al., 2008, 2019; Kottelenberg and Lehrer, 2013). It is important to note that the Quebec program provided cheap but rather low-quality childcare. In contrast, the German or Scandinavian daycare system are considered to deliver mediocre and high-quality care, respectively.[5] On the other hand, there are high-quality programs specifically targeted at disadvantaged families such as the Perry Preschool Program, the Abecedarian Project or the Head Start Program. All programs were shown to positively affect health (e.g. Carneiro and Ginja, 2014; Conti et al., 2016; Heckman et al., 2010). Furthermore, the daycare expansion in Germany during the 1990s for children aged three and older was shown to positively affect children's physical health (decline in physician recommendations for compensatory sport (Cornelissen et al., 2018) and fewer weight problems and better performance in the gross motor skills test (Lauber, 2015)). All of the presented studies rely on survey data. In contrast, van den Berg and Siflinger (2020) use detailed administrative health records for one region in Sweden and exploit a daycare reform in Sweden that increased daycare exposure by reducing fees of public day care. Their results suggest that daycare attendance leads to an improvement in mental health at primary school age and to a substitution of infections from primary school ages to low ages. Informal daycare arrangements serve as the counterfactual to public daycare in this setting.[6]

In this paper, I consider both physical and mental health outcomes as well as healthcare consumption and costs. Specifically, in terms of physical health, I analyze three sets of communicable diseases, namely the ICD-10 groups of *infections* (A00-B99), *respiratory diseases* (J00-J99) and *ear diseases* (H60-H95), and *obesity* (E65-E68) as a non-communicable disease. For mental health including non-cognitive skills, I consider the ICD-10 group of *mental and behavioral disorders* (F00-F99).[7] To measure healthcare consumption I assess the annual number of treatment cases. The healthcare costs measure includes all costs billed from ambulatory care doctors within one calendar year.

[5] Also Zerpa (2018) finds evidence for negative health impacts (increase in sick days) of State Pre-K programs in the US. However, these effects seem to fade out when children enter elementary school.

[6] Aalto et al. (2019) study the health effects of daycare on children with unemployed parents using registry data on in-patient care and prescription drugs in Sweden. They do not find an effect on hospitalization rates for children aged 2–3 years. However, the hospitalization rate due to infections increases for preschool aged children, while there is no effect on the overall hospitalization rate. Furthermore, they provide evidence for the hygiene hypothesis (explained on p.5) due to reduced prescriptions for allergies and asthma at elementary school age.

[7] Note, these sets of conditions correspond to the 2-digit-level ICD-10 codes. I also provide results for more detailed diagnoses within these sets of conditions (3- and 4-digit ICD-10 codes).

Ex-ante there is no clear prediction of the direction of the effects as daycare attendance may affect these outcomes through several channels: First, communicable diseases such as infections, respiratory conditions and ear problems are very prevalent among (young) children and anecdotal evidence suggests that children who just entered daycare or school are very frequently affected. This observation is in line with the so-called hygiene hypothesis stating that exposure to viruses and bacteria at early ages initiates an immunization process that leads to more infections in the short-run but fewer infections at older ages (Strachan, 1989)[8]. In fact, evidence from the medical and epidemiological literature suggests that there is an association between daycare attendance (at young ages) and the prevalence of communicable diseases (e.g. de Hoog et al., 2014; Kamper-Jørgensen et al., 2006; Watamura et al., 2010).[9] On the one hand, a substitution of infections from elementary school age to daycare age might be desirable from a welfare point of view as sickness absence of students in schools may be reduced. On the other hand, some diseases, e.g. acute respiratory infections, might be particularly dangerous for very young children (e.g. Kamper-Jørgensen et al., 2006) and might lead to more hospitalizations and antibiotic prescriptions. Higher antibiotic intake in children may have also adverse long-term effects on the cognitive development and other health outcomes such as obesity (e.g. Baron et al., 2020; Mbakwa et al., 2016). Furthermore, sickness absence at daycare centers could disrupt the relationship to daycare teachers and fellow children in daycare and thus be harmful to children's early development.

Second, care actors may play an important role in children forming health habits (e.g. through movement habits and nutrition). Health habits are formed early in life, thus childhood obesity is strongly correlated with adult overweight. Being overweight or obese is an important determinant of future health problems and chronic conditions (such as cardiovascular diseases and diabetes (e.g. Must et al., 1999)). Interventions to prevent obesity were shown to be particularly effective in children below the age of six (Davis and Christoffel, 1994; Waters et al., 2011). Additionally, eating habits – that are crucial causes of obesity – are likely developed early in life (e.g. Birch, 1999). Therefore, daycare attendance has the potential to influence health habits and thereby preventing obesity.

[8]Originally, the hygiene hypothesis was developed as an explanation for a reduction in hay fever and asthma diagnoses for children with many siblings as they are exposed to many microbial compounds early in life. Later, the hygiene hypothesis has also been related to a more general immunization process, not only affecting allergic illness but also other inflammatory diseases (e.g. Briggs et al., 2016; Oikonomopoulou et al., 2013; Schaub et al., 2006)

[9]Note that these studies do not take the endogeneity between daycare attendance and health into account and thus do not provide causal evidence.

Third, there is evidence that daycare attendance is associated with the development of non-cognitive skills (e.g. Baker et al., 2008; Felfe and Lalive, 2018) and that the formation of non-cognitive/social skills is at least equally as important as the development of cognitive skills. Cunha and Heckman (2007) for example find that non-cognitive skills promote the formation of cognitive skills but not vice versa. Similarly, Heckman (2006) shows that non-cognitive skills are as important for school enrollment decisions as cognitive skills. Currie and Stabile (2006) point out that mental disorders have larger adverse effects on future reading and mathematics test scores than physical health problems. A severe measure of difficulties in non-cognitive/social development are diagnosed mental and behavioral disorders. Thus, I focus on below-average non-cognitive/social development rather than capturing the full range of non-cognitive/social development of children. Furthermore, there is evidence that the child care setting can affect the salivary cortisol level in young children. Higher cortisol levels can be evidence for stress and decrease the antibody levels which can result in greater illness frequency (Watamura et al., 2010). This underlines that physical and mental health are interrelated and emphasizes the importance to study both dimensions to capture the overall impact of daycare attendance on children's health.

Lastly, not only daycare attendance itself but also changes in the environment of the child due to the expansion may affect their health. For example, there might be some kind of health surveillance at daycare centers that track children's health (e.g. traces of abuse). This could result in more frequent doctor visits and thereby more diagnoses. Additionally, as Müller and Wrohlich (2020) have shown, the daycare expansion has increased female labor market participation. On the one hand, employed parents need a doctor's note to take sick leave when their child is sick. Therefore, they might take their child to the doctor with less severe conditions resulting in more diagnoses than parents who care for their child at home. On the other hand, employed parents have more time pressure and might therefore skip doctor visits for less severe conditions resulting in fewer diagnoses. Increased maternal labor market participation might further potentially lead to a positive income effect.[10] As income and health are positively correlated, this could lead to better health of children who were affected by the reform. Based on these contradicting predictions, it remains an empirical question whether the expansion affected children's health outcomes and whether they increased or decreased in certain age groups.

[10] This has not been studied for the daycare expansion yet. However Nicoletti et al. (2020) show, that the direct negative effects on mid-childhood and teenage outcomes induced by increasing mothers' labor supply is fully compensated by increased income.

My results provide evidence that early daycare attendance increases the prevalence of respiratory and infectious diseases at age one to two but decreases the prevalence at older ages. The reductions are most pronounced at the first years of primary school (age six to eight). In line with the hygiene hypothesis, this suggests a substitution of illness spells from the first years of elementary school to the first years of daycare. The intention to treat effect amounts to about 1–2% fewer diagnoses for respiratory and infectious diseases (depending on the age group), while the treatment-on-the-treated effect is about 5%. The increase in infections at age 1–2 years corresponds roughly to the decrease at elementary school age, while the increase in respiratory diseases at early ages is about twice as large as the decrease at older ages. For mental health and obesity, I find no significant changes. Health care consumption increases at ages 1–2 years while it decreases at ages 3–5 years. There is no effect on healthcare consumption for older children. Despite changes in the prevalence of diagnoses and number of doctor visits, there is no effect on healthcare costs for any age group. The findings are robust to a large set of robustness checks such as different definitions of the treatment status and the expansion period and the application of randomization inference methods to obtain valid p-values.

These results raise the question whether the substitution of illness spells for infections and respiratory diseases from the first years of elementary school to the first years of daycare is beneficial or not. In terms of healthcare costs arising in the first ten years of life, the daycare expansion appears to be neither beneficial nor costly. However, to evaluate the welfare effects in terms of health of the reform other aspects such as severity/duration of illness spells at different ages, sickness absence at school/daycare, long-term health effects and opportunity costs for parents need to be considered. I will carefully evaluate these aspects for a "back-of-the-envelope"-calculation of the welfare effects in terms of health of the reform as this project proceeds.

This paper is structured as follows. In the next section, I describe the institutional setting, particularly the daycare expansion. In Section 3, I present my data with focus on the construction of my sample and the outcome variables of interest. Next, I outline my empirical strategy and discuss the underlying assumptions. In Section 5 I present my empirical results, discuss them and provide a heterogeneity analysis as well as an extensive set of robustness checks. Section 6 concludes.

2 Institutional Setting

In West-Germany, traditionally female labor market participation of mothers with young children is low (e.g. 35% in 2005 for mothers with children below the age of three (Müller and Wrohlich, 2020)).[11] Besides incentives set by the tax and transfer system, one frequently quoted reason is the low supply of formal care arrangements for (young) children. In 1996, the child and adolescent support law ("Kinder- und Jugendhilfegesetz"), that guaranteed every child aged three and older a kindergarten place, was implemented. Today, almost all children visit a daycare center for at least one year before entering school. Other policy reforms affecting the supply of formal child care slots have only been initiated since the middle of the 2000s (Spieß, 2011). In 2005, the daycare expansion law ("Tagesbetreuungsausbaugesetz", TaG) was passed aiming at expanding formal care slots for children under the age of three (230,0000 additional slots in West-Germany), particularly focusing on care offered by childminders ("Tagespflege"). In 2007, a summit ("Krippengipfel") of the federal government, the federal states, and the counties reinforced the aim of the 2005 mandate and set the target of a 35% daycare coverage rate for children under three years by 2013. Finally, in 2008 the law on support for children ("Kinderförderungsgesetz") was introduced that commits states to a gradual expansion of daycare supply for children below three years. The law also entailed a legal entitlement to every parent with a child aged one to three years to a subsidized daycare slot by August 2013.

These reforms induced a large expansion of publicly subsidized daycare slots in both, West and East Germany. However, the expansions in East and West Germany differed in their extent and the starting level. In West, Germany daycare coverage for children under three years increased from about 12% in 2008 to 29% in 2018, while it increased in East Germany from about 43% to 55% during the same period (Destatis, 2018*b*).[12] I restrict the analysis to West Germany[13] as the situation in West and East Germany is very different and difficult to compare. The development of daycare coverage in West-Germany between 1994 and 2018 is depicted in Figure 1. It becomes visible that daycare coverage for children under three years was very low (below 5%) up until the early 2000's. From the mid 2000's West-

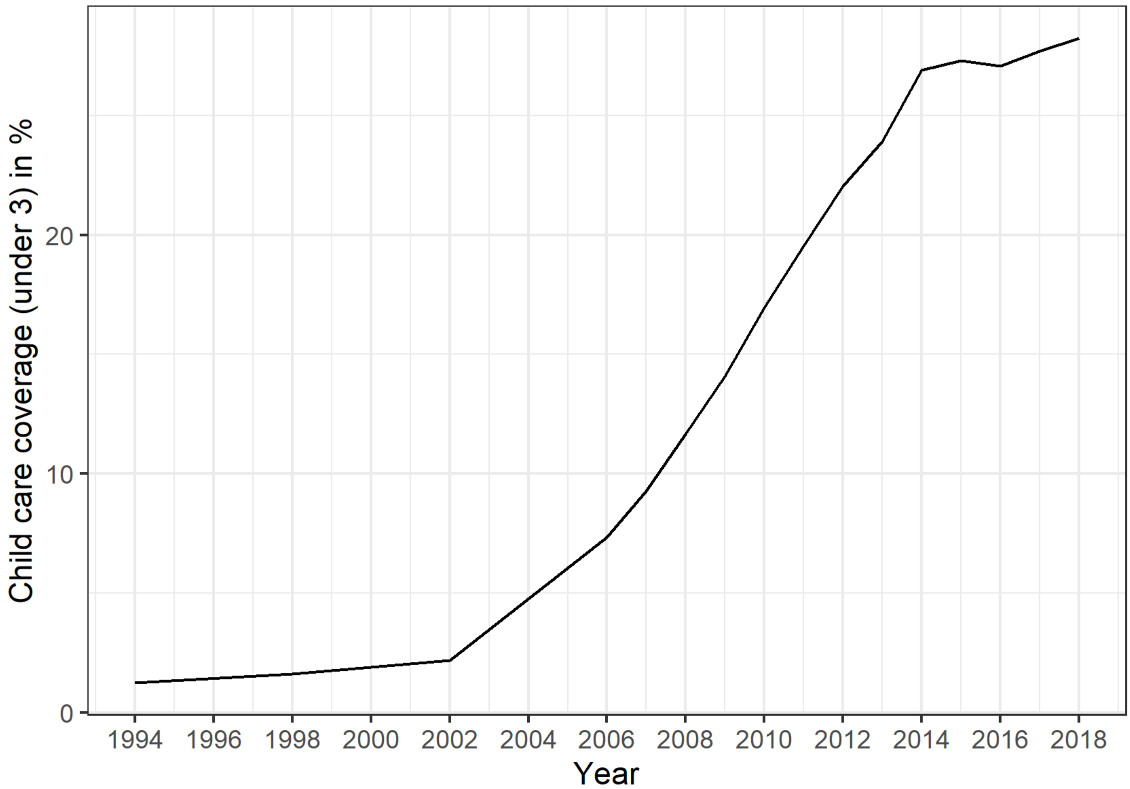
[11] Due to the division of Germany, social norms as well as family policies have developed differently in East- and West-Germany. Female labor market participation as well as daycare coverage is still today much higher in East-Germany compared to West-Germany (e.g. Müller and Wrohlich, 2020).

[12] Daycare coverage is defined as the share of children being in daycare. As there has been and still persists an excess demand for daycare slots, I assume a full take-up of newly created daycare slots for children below three years in the subsequent analyses (Müller and Wrohlich, 2016; Wrohlich, 2008).

[13] Berlin is excluded from the analysis.

Germany experienced a steep increase that ran flatter from 2014 onward. The increase experienced up until 2018 was large, however, the goal of 35% coverage was not reached. Furthermore, the expansion created large regional variation in the expansion speed. Figure A.1 in the Appendix shows that first, in 2008 the majority of West-German counties had a coverage rate below 20%. In 2018, the majority of counties lies above 20%, many above 30%. Secondly, is becomes apparent that the expansion speed differed substantially across counties.

Figure 1: Daycare coverage U3 West-Germany



Note: The graph shows the daycare coverage rate for children below three years in West-Germany.
Source: Destatis 1994-2018, own calculations

Germany is characterized by a publicly subsidized daycare system. Public daycare is provided by communities while private providers that are mostly publicly subsidized include religious non-profit, non-religious non-profit, or commercial institutions (Muehler, 2010). Overall, 98% of providers are considered non-profit providers (Destatis, 2018*a*). Generally, daycare is highly subsidized by the the federal government, the states and the municipalities, but the exact amount and source of funds depends on state-specific contracts. Across states, on average the main funding source of non-profit providers are public funds (cover 75-80% of costs), 10% of the costs is covered by the providers themselves and the remainder by parents (Spiess, 2008). Compared to other OECD countries, Germany's public expenses compared to the GDP are only slightly above average. Daycare fees typically range between 5 to 9% of net family income (Schmitz et al., 2017), which is below the OECD average (OECD, 2019*a*).[14] The general objective, strategy and funding of daycare is determined at the federal level. However, operational planning and the implementation of objectives is managed by municipal governments and/or youth welfare offices ("Jugendämter"). Thus, the structure and organization of daycare vary between states (Hüsken, 2011; Müller and Wrohlich, 2020). Local authorities estimate the local demand for daycare slots and develop an expansion strategy accordingly. The procedure is not uniform across municipalities and thus leads to the observed differences in the expansion speed.

Previous work has established that the reform increased female labor market participation (Müller and Wrohlich, 2020), reduced child maltreatment (Sandner, 2019) and improved children's socio-emotional skills (Felfe and Lalive, 2018).

3 Data

For the analysis, I use administrative data covering the years 2009-2019[15], collected by all public health insurance funds in Germany. The data are based on the database of claims of all publicly insured individuals in Germany as collected by the Association of Statutory Health Insurance Physicians and then forwarded to the National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung, KBV). In the data, physicians record a standardized diagnosis for each claim in order to be reimbursed by the health insurance. In Germany, health insurance is mandatory and characterized by a public insurance system and a private insurance system. Nearly 90% of the German population is covered by one of the public health insurance funds.[16] Only individuals with earnings exceeding a certain threshold[17] and individuals in specific occupational groups (e.g., civil servants and self-employed) are allowed to opt out of the public system and to sign up with a private insurance company instead. Children are covered by the health

[14]For a more detailed overview of the organization and funding of German daycare centers please refer to e.g. Huebener et al. (2020).

[15]2020 is excluded from the analysis due to the Covid-19 pandemic.

[16]Public health insurance is financed primarily through mandatory contributions from employers and employees, along with tax revenues. Contributions are pooled in the Central Health Fund (Gesundheitsfonds) and reallocated to the sickness funds according to a morbidity-based risk adjustment scheme. There are currently about 109 health insurance funds. For more information about the German health insurance system, see OECD (2019*b*).

[17]The income threshold for 2020 was 62,500 euro ($\approx 74,500$ dollar) per year.

insurance of their parents without extra fees being due.

3.1 Sample

I have access to data covering 2009 through 2019, thus I am able to consistently estimate age-specific health effects. To do so, I split the sample into four different age groups: toddlers (1–2 years), kindergarten aged children (3–5 years), early elementary school aged children (6–8 years) and older elementary school aged children (9–10 years). Exploiting the daycare expansion described in Section 2, I construct a treatment and a control group depending on the birth year of the children and their county of residence. In the analysis, I focus on children born between 1999–2015. Comparing the number of children that are publicly health insured (Gesundheitsministerium, 2020) with official birth records for each birth cohort (Statista, 2021) suggests that I cover about 86% of all children born in Germany in the respective birth years.[18]

The data include information about all diagnoses patients received during the observed period. Each diagnosis constitutes a new entry meaning that the number of observations equals the number of diagnoses over the observed time period. Thus, the sample is unbalanced as patients only appear if they received outpatient care including a diagnosis. Based on this information, I construct a balanced sample with yearly information for all publicly insured children.[19] The final data set includes about 550,000–650,000 children per birth

[18]Note, there is only aggregated data on the number of publicly health insured individuals for 0–14 year old children available for 2004–2020. To get an estimate of how many children are covered in the data, I add up the official births for the respective birth cohorts that are 0–14 years old for each year between 2009–2019. The average share of officially born children with public health insurance amounts for all years to about 86% (Gesundheitsministerium, 2020; Statista, 2021).

[19]Analogous to Barschkett et al. (2021), first, I create variables indicating the number of times an outcome, for example respiratory conditions, was diagnosed in a specific period. Secondly, I aggregate the data to a yearly level such that each patient appears only once per year. Finally, I balance the data by imputing information for patients without outpatient care in a specific year. By definition, all outcome variables are zero as the patient did not receive a relevant diagnosis during this year. The definition of my outcome variables is analogous to van den Berg and Siflinger (2020) and Barschkett et al. (2021).

cohort resulting in about 11 million children overall[20][21].

3.2 Outcome variables

I define measures for physical and mental health using ICD–10 codes. Instead of estimating the effect for about 70,000 different diagnoses categorized by the ICD-10 codes, I use broader 2–digit and 3–digit categories. As described before, I aggregate them on year level and compute the annual number of diagnoses per child. As an additional measure that captures the extensive margin, I construct binary indicator variables, that mark whether a child had a specific diagnosis at least once in a given year.

Similar to van den Berg and Siflinger (2020), I assess three different aspects of health: physical health (communicable and non-communicable diseases), mental health and health care consumption. My core physical health measures for communicable diseases capture the following three sets of conditions: *respiratory diseases*, *infections* and *ear problems*. *Respiratory diseases* are categorized in ICD–10 codes J00–J99. In Germany, the most frequent diagnosis code of all ICD-10 codes used by pediatricians is Acute upper respiratory infections of multiple and unspecified sites (J06). Other respiratory diseases that are also among the top 50 diagnoses used by pediatricians are Asthma (J45), Acute bronchitis (J20), Vasomotor and allergic rhinitis (J30), Acute tonsillitis (J03), Other respiratory disorders (J98), Acute nasopharyngitis (common cold, J00), Bronchitis, not specified as acute or chronic (J40), Acute pharyngitis (J02) and Acute laryngitis, tracheitis (J04) and Chronic rhinitis, nasopharyngitis and pharyngitis (J31) (ZI, 2015). In addition to the analysis of the effects on the aggregated condition *respiratory diseases*, I will also investigate the mentioned subgroups. *Infectious diseases* are comprised in ICD-10 codes A00–B99. Generally infections include any bacterial or viral infection. Common infections occurring in childhood are Other infectious diseases (B99), Viral infection of unspecified site (B34) and Other gas-

[20]Children who did not receive any outpatient care during the 11 year observation period are not included in my sample. However, Kamtsiuris et al. (2007) state, that 95% of 0–2 year old children visit a pediatrician at least once per year. Additionally, more than 90% of children make use of the individual early diagnostic tests. Thus, given that I observe individuals over 11 years, the share of children not receiving any outpatient care should be negligible.

[21]In the dataset, children are identified via a unique patient ID that is based on name, first name and date of birth. Note, I only observe the ID, not the underlying information. Due to errors in recording name, first name and birth date throughout the billing process, some patients have multiple IDs, i.e. they have one correct ID and then other IDs that were created due to an error in the spelling of the name or date of birth. The majority of these "wrong" IDs only appear once or twice during the observation period. These errors should not be systematically and thereby do not pose threats to my identification strategy. The current analysis is conducted based on all IDs. I am currently working on a solution to be able to match "wrong" IDs with their correct ID.

troenteritis and colitis of infectious and unspecified origin (A09) (ZI, 2015). Also here, I will analyze the subgroups in addition to the aggregated condition *infections*. *Ear problems* are represented by the ICD–10 codes H60–H95. These codes include diagnoses on the external ear, the middle ear and the internal ear. Here the most common diagnoses occuring among children are Suppurative and unspecified otitis media (H66), Other hearing loss (H91) and Nonsuppurative otitis media (H65) (ZI, 2015). The three sets of conditions are mutually exclusive meaning that the pediatrician (or another healthcare professional) settles on one ICD-10 code as a diagnosis. However, the conditions are closely related and could could be in a causal relationship to each other. In particular, many diagnoses concern contagious diseases that are common in childhood and often transmit between children; thus they likely also spread in daycare centers. Many infections may go along with respiratory problems and could cause subsequent ear problems. Furthermore, some respiratory diseases or ear problems concerning inflammations could result from infections. Thus, children could have multiple of these conditions per year.

As an additional measure for physical health (non-communicable disease), I assess obesity. Obesity is a very common condition among children and adolescents in many industrialized countries including Germany. In Germany, 15.4% of all children between 3 and 17 years count as overweight, while almost 6% are defined as obese (Schienkiewitz et al., 2018). Obesity has its onset often early in childhood[22] and is influenced by health behavior and general lifestyle. There might be differences for children in different care types (e.g. home care vs. daycare). Obesity is categorized in ICD-10 codes E65–E68.

To assess the effects on *mental health* including behavioral problems, non-cognitive abilities, and mental health problems, I analyze the effect on ICD–10 codes F00-F99. The most frequent mental health diagnoses among children in Germany are Specific developmental disorders of speech and language (F80), Hyperkinetic disorders (F90), Unspecified disorder of psychological development (F89), Specific developmental disorder of motor function (F82), Other behavioural and emotional disorders with onset usually occurring in childhood and adolescence (F98) and Mixed specific developmental disorders (F83) ZI (2015). Note, I measure non-cognitive skills by diagnosed mental and behavioral disorders, which is certainly a more extreme measure than typical measures obtained from survey data (e.g. SDQ-Index). Thus, I rather focus on below-average non-cognitive developments than cap-

[22] At age 3 to 6 about 10.8% of girls and 7.3% of boys are overweight (Schienkiewitz et al., 2018)

turing the full range of non-cognitive abilities of children. Measures in survey data, however, might under report problems in the cognitive development as survey respondents (mostly parents) are first not professionally trained to recognize behavioral disorders and secondly might have difficulties to accept that their child exhibits behavioral disorders.

Lastly, I consider health care consumption and healthcare costs. Healthcare consumption is defined as doctor visits which are measured as treatment cases, aggregated at the calendar year level (official term: "Artzfälle"). One treatment case is defined as a treatment of an insured person by a doctor in a quarter, billed to one public health insurance fund.[23] Thus, if a child visits two different doctors in a quarter, she has two treatment cases in that specific quarter.[24] I aggregate quarterly cases to the calendar year level, thus counting the number of quarterly doctor cases per year. This means that a patient who visits every quarter only one and the same doctor would have a yearly count of four doctor cases, irrespective of the actual number of visits to this doctor per quarter. Similarly, healthcare costs are documented on the quarter level and include all costs billed from ambulatory care doctors. I also aggregate the costs to the calendar year level and adjust them to 2009 fees.[25]

Summary statistics for the outcome variables including the sample means for annual diagnoses as well as the prevalence of the diseases are shown in Table 1. For all three sets of communicable diseases, the number of diagnoses per year as well as the prevalence decreases with age. On average, 1–2 year old children have about 1.4 infections per year while 9–10 year old children have only 0.7 infections per year. Respiratory diseases have the highest prevalence among all considered outcomes, e.g. 80% of 1–2 year olds have a respiratory diagnoses at least once per year. In contrast, the likelihood of mental disorders and obesity increases with age. The annual number of treatment cases decreases with age, 1–2 year olds have on average 6.3 treatment cases per year, while 9–10 year olds have only 4.9. In line with decreasing treatment cases with age, also healthcare costs are higher for younger children (on average 320 Euro per year for 1–2 year old children) than for older children (249.23 Euros per year for 9–10 year old children).

[23] Since treatment cases are recorded this way in the data, I do not have the possibility to define the variable differently for my application.

[24] If she visits only one doctor but switches health insurance providers, she would also be assigned two doctor visits. However, since only 3% of children in my sample switch health insurance providers, this issue is negligible.

[25] Fees are adjusted to 2009 fees. This adjustment accounts for the general increase in the fee level and specific changes to the medical system (The time series "Honorarumsatz je Behandlungsfall in Euro" from 2009–2018 was used to adjust fees (KBV, 2019)).

Table 1: Outcomes

	1-2 years	3-5 years	6-8 years	9-10 years
Infections (no. per year)	1.39 (1.59)	1 (1.29)	0.78 (1.11)	0.66 (1.04)
Infections (prevalence)	0.63 (0.48)	0.53 (0.50)	0.46 (0.50)	0.41 (0.49)
Ear diseases (no. per years)	0.58 (1.13)	0.84 (1.47)	0.45 (1.09)	0.28 (0.84)
Ear diseases (prevalence)	0.33 (0.47)	0.39 (0.49)	0.24 (0.43)	0.16 (0.37)
Respiratory diseases (no. per year)	2.85 (2.67)	2.65 (2.75)	1.85 (2.35)	1.58 (2.22)
Respiratory diseases (prevalence)	0.81 (0.39)	0.77 (0.42)	0.65 (0.48)	0.58 (0.49)
Mental disorders (no. per year)	0.31 (0.87)	0.87 (1.63)	1.057 (2.15)	1.03 (2.38)
Mental disorders (prevalence)	0.18 (0.38)	0.37 (0.48)	0.33 (0.47)	0.27 (0.45)
Obesity (no. per year)	0.02 (0.21)	0.04 (0.32)	0.06 (0.40)	0.10 (0.50)
Obesity (prevalence)	0.01 (0.12)	0.02 (0.15)	0.03 (0.18)	0.05 (0.22)
Treatment cases	6.33 (3.84)	6.14 (4.04)	5.28 (7.46)	4.92 (8.91)
Healthcare costs	320.0 (312.57)	287.40 (319.93)	244.57 (393.18)	249.23 (450.29)
Observations	9,042,454	16,840,400	17,167,518	11,674,867

Note: Reported are means and standard deviations in parentheses. No. per year indicates count variables, i.e. contains the number of diagnoses per year. Prevalence indicates dummy variables, i.e. indicates children that had at least one diagnosis per year. Costs are fee-adjusted. *Source:* KBV, own calculations

3.3 Control variables and daycare coverage rates

The KBV data only includes few individual-level socio-demographic characteristics, such as age, gender, year, birth month and county of residence. Additionally, county-level ("Landkreise") information such as unemployment rate, GDP per capita, household income and education are available and can be used for heterogeneity analyses. The data does not contain information about individual childcare arrangements, i.e. I do not observe if children attend daycare. Therefore, I merge the KBV data with county-level information on the share of children enrolled in daycare.[26] After multiple county reforms that reorganized the counties, the most recent number of counties in Germany amounts to 401. Since 2006 data on daycare coverage is provided by the German Statistical Office. Before 2006, data for 1994, 1998 and 2002 were gathered by German Youth Institute based on material by the Statistical Offices of federal states. In the main analysis, I focus on the main expansion period of daycare slots (2008–2012). In additional analyses I also make use of earlier and later years. I restrict the analysis to West Germany[27] (323 counties) as East Germany traditionally has higher daycare coverage rates for very young children. Therefore, the situation in West and East Germany are very different and difficult to compare.

[26]There is an excess demand for childcare in all regions of Germany (see, e.g. Müller and Wrohlich (2016)). Therefore, I argue that this indicator can be interpreted as the supply of daycare slots.

[27]Berlin is excluded from the analysis.

4 Empirical strategy

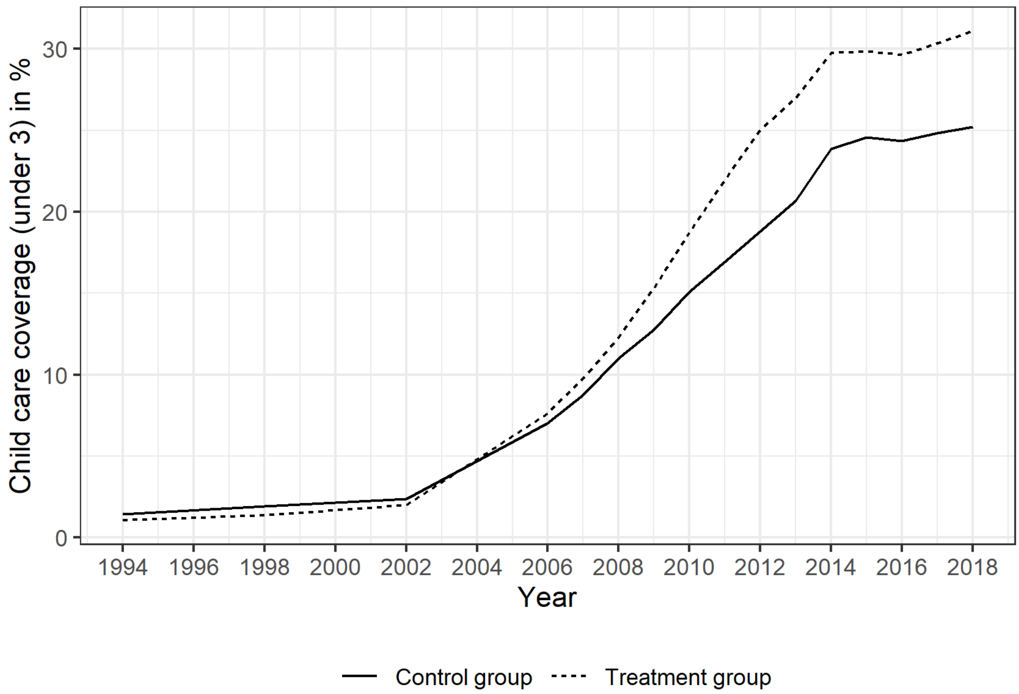
To estimate the effect of the daycare expansion on children's health outcomes, I exploit spatial and temporal variation in the daycare expansion by employing difference-in-difference (DiD) as well as event study (ES) approaches. Below, I am going to describe my main empirical strategy (DiD, ES and generalized DiD) and discuss potential threats to identification and alternative specifications to validate my results.

4.1 Difference-in-differences approach

In my DiD specification I follow Havnes and Mogstad (2011) and Müller and Wrohlich (2020) by comparing health outcomes of children at different ages for 1–2 year olds before and after the expansion from counties where daycare expanded a lot (treatment group) and counties with little or no increase in daycare coverage (control group). The daycare expansion started in the mid 2000's, had its peak between 2008 and 2012 and substantially slowed down from 2014 onward (see Figure 1 in Section 2). Based on this, I define 2008–2012 as the main expansion period. Setting 2008 as the starting year gives municipalities some time to adjust to the 2007-announcement of a legal entitlement to a daycare slot for all children aged one year and older from 2013 onward. Furthermore, 2008–2012 was the period with the largest growth in daycare coverage. Therefore, post-reform cohorts born 2007–2011 were affected with full force, whereas the phase-in cohorts born 2005–2006 were affected to a lesser extent. All cohorts born before 2005 were not affected by the expansion (pre-reform cohorts). In robustness checks, I use different definitions of the main expansion period to ensure that my results are robust to changes in the exact choice of the expansion period.

To divide counties into the treatment and control group, I order countries according to the percentage point increase in daycare coverage rates from 2008–2012. I then separate the sample at the median, the upper half constituting the treatment counties and the lower half the control group. Figure 2 depicts daycare coverage rates before, during and after the expansion in treatment and control counties. The graphs move almost in parallel up until circa 2008, while treatment counties experience a steeper increase in daycare coverage from 2008 onward. Thus, I compare counties that distinctly differ in their expansion speed within the main expansion period. In robustness checks, I provide evidence that my results are robust to changes in the definition of the treatment group by choosing other cutoffs than the median (e.g. upper 30% vs. bottom 30%).

Figure 2: Daycare expansion during 1994 and 2018



Note: The graph shows the daycare coverage rate for children below three years in West-Germany comparing treatment and control counties.
Source: Destatis 1994-2018, own calculations

My main regression model, estimated by OLS, can be defined as

$$Y_{it} = \psi_t + \gamma_1 Treat_i + \gamma_2(Treat_i \times Phasein_t) + \theta(Treat_i \times Post_t) + X_{it}\beta + \varepsilon_{it} \quad (1)$$

where Y_{it} represents the health outcomes of child i born in year t. ψ_t are year fixed effects and X_{it} is a vector of control variables that includes age, gender and county fixed effects. $Treat_i$ is a dummy variable that indicates whether child i lived in a treatment county, $Phasein_t$ is 1 if child i was born in year $t \in [2005, 2006]$[28] and $Post_t$ turns 1 if child i was born in year $t \in [2007, 2011]$. Interacting $Treat_i$ and $Post_t$ marks all children that were affected by the expansion, i.e. children that were born between 2007–2011 and lived in a treatment county. Thus, θ is the coefficient of interest and captures the intention-to-treat (ITT) effect of the expansion. I interpret this as an ITT effect, as my model estimates the reduced form impacts on all children from post-reform cohorts who reside in the treatment

[28]Note, the $Phase - in$ dummy is excluded in the analysis for 3–5 year old children, as data is only available from 2006 for 3 year old children.

area. The benefit of estimating an ITT effect is, that it captures the full reform impact. Thus not only the effect on treated children is portrayed but also potential spill-over effects on for example siblings that were themselves not affected by the reform, peer effects on other children who were not attending daycare and changes in both formal and informal care arrangements (e.g. grandparental care) are captured. However, the ITT averages the effect over all children in treated counties. Therefore the size of the effect is difficult to interpret and needs to be weighed against the size of the expansion. In order to do so, I compute the treatment-on-the-treated (TT) effect by scaling the ITT with the first stage results. In the first stage, I estimate the same model as in Equation 1 with the daycare coverage rate in county c for birth cohort t on the left hand-side. θ gives then the change in the daycare coverage rate for affected counties. I arrive at the TT by calculating TT = ITT/first stage. The TT represents the effect of daycare exposure (per daycare spot) on children born in post-reform cohorts who live in the treatment area. In my main results, I report both the ITT and the TT estimates.

To validly estimate the effect of the daycare expansion within a DiD approach, two important factors must be considered. First, the common trend assumption needs to hold, i.e. in absence of the expansion, treatment and control counties should have evolved in the same way. This assumption is addressed in Section 4.2. The second issue involves the correct calculation of standard errors. In order to correct for possible serial correlation of the error terms, I report heteroskedasticity-consistent standard errors that are clustered at county-level. This leads to asymptotically valid inference, however, in finite samples such as mine (323 clusters) the problem is still present (Cameron and Trivedi, 2005). The number of clusters in my study is not as small as compared to studies on the state level (e.g. Pfeifer et al., 2020), but the counties (clusters) do differ in size and outcomes potentially covary with cluster size. Therefore, I apply randomization inference (RI) to obtain valid p-values (Gerber and Green, 2012, Ch.3). The intuition behind RI is the following: In order to overcome sampling uncertainty, the distribution of the treatment variable is repeatedly randomized. As with large sample sizes the number of possible randomizations is too large to handle, one can sample at random from the set of all possible random assignments to approximate the sampling distribution. To do so, first the test statistic from the original data set is computed. Second, the treatment is randomly assigned and the test statistic is computed again. Third, the previous step is repeated, in my case 1000 times. Lastly, the generated distribution is compared to the original test statistic and it is evaluated with what probability the null hypothesis can be rejected. Thus, this procedure creates an exact reference distribution as it fully describes the sampling distribution under the null hypothesis. My effects can be evaluated against this distribution which yields more precise p-values than comparing them to standard approximations such as t-distributions. In many applications, the randomization distribution has fatter tails than the standard t-distribution. Therefore, evaluating the effects against the RI distribution is more conservative (Gerber and Green, 2012; Pfeifer et al., 2020).[29] In my application, children's treatment status depends on two variables: their county of residence and their birth cohort. However, before applying RI, I interact the two variables and then permute the generated treatment status.

The information of the county of residence of the children is essential to identify the treatment effect. From 2009 onward, I observe child i's county of residence annually. However, for all children born before 2009, I do not have information about their county of residence prior to 2009. Therefore, I approximate their county of residence by assuming that they live in the county of the first observation and define treatment and control group accordingly. To be consistent, I use the county of residence of the first observation for all children (2009 for all children born before 2009 and the respective birth year for all children born after 2009). Thus, in my main specification, I ignore if children move during the observation period. I deem this assumption plausible since only 7% of the children in my sample move during the observation whereof only 2% of children move from a treatment to a control county or vice versa. In robustness checks, I provide evidence that my results are robust towards this approximation. To verify this, I use other assumptions (e.g. use the last observed county of residence instead of the first observation) and I exclude all children that moved during the observation period from the analysis.

Because I control for county-specific fixed effects, it is not necessary that the daycare expansion is unrelated to county characteristics. It is useful, however, to understand the determinants of the expansion across counties. In Table B.1, I investigate differences in socio-demographic county characteristics between treatment and control counties in 2008. For most characteristics, I can depict only minor differences between treatment and control counties. Interestingly, at the beginning of the expansion period, treatment counties already depicted a slightly higher daycare coverage rate (12.3 vs. 11%). Furthermore, the unemployment rate was slightly higher in control counties (5.8 vs. 6.3%), the share of migrants

[29] For more details about the origin of randomization inference, the procedure and examples of applications see e.g. Gerber and Green (2012) or Pfeifer et al. (2020)

was higher in control counties (6.7 vs. 10.0%), population density was almost twice as large in control counties compared to treatment counties (357.5 vs. 760.8) and GDP per capita was also higher in control counties (28,061.3 vs. 33,285.9 Euros). Thus, overall treatment and control counties are fairly comparable in their socio-demographic composition, with share of migrants and population density being notable exceptions.

4.2 Event Study approach

The DiD approach controls for unobserved differences between children from treatment and control counties as well as between children born in different years. The identifying assumption, namely the variance-weighted common trend assumption, states that in absence of the expansion, the change in the outcomes of interest for the 1–2 year olds before and after the expansion would have been the same in the treatment and the control counties. Thus, if this assumption does not hold, the effects cannot be interpreted causally. The validity of this assumption cannot be tested directly. However, I perform several checks on the plausibility of this assumption. First, I exclude year and county fixed effects as well as other control variables to check the sensitivity of the results towards the inclusion of these variables. Second, I implement a placebo regression employing a chronic disease (diabetes mellitus) that should be unaffected by exposure to daycare or other environmental factors as an outcome variable. Third, I provide event study graphs based on similar regressions as presented in Equation 1. Specifically, I estimate the following specification

$$Y_{it} = \psi_t + \theta(Treat_i \times Cohort_i) + X_{it}\beta + \varepsilon_{it} \tag{2}$$

where $Cohort_i$ represents the birth year of child i. All other variables are the same as in Equation 1. If no pre-trends are present, the coefficients on the interaction between $Cohort_i$ and $Treat_i$ should be small and insignificant for all birth cohorts born prior to the expansion. The identification in an event study approach is robust to time-varying treatment effects (Goodman-Bacon, 2021). Graphical evidence is shown in Section 5.

4.3 Generalized DiD

The DiD estimator captures the average causal effect of additional daycare slots following the expansion in treatment relative to control counties. When consulting Figure A.1 it becomes evident that treatment intensity does not only vary between treatment and control counties but also within the two groups. Apart from varying the treatment definition in robustness checks, I also estimate a generalized DiD specification where I regress the outcomes directly on the daycare coverage rate in each county controlling for year and county fixed effects as well as the same set of controls as in the standard DiD framework. With this approach I closely follow Müller and Wrohlich (2020).

Specifically, I estimate

$$Y_{it} = \psi_t + \theta cc_{ct} + X_{it}\beta + \varepsilon_{it} \tag{3}$$

where cc_{ct} is the average child care coverage rate in county c at time $t \in age[1,2]$.[30] The remaining variables are the same as in Equation 1. Under the common trends assumption and the assumption that the marginal effect of an additional daycare slots is constant, θ can be interpreted as the causal effect of an increase in the daycare coverage rate on the the outcomes of interest.

As the KBV data are only available from 2009 onward, the standard DiD approach described in Section 4.1 can only be applied for children aged 3 years and older. The reason for this is that for younger children a pre-period is missing, as only birth cohorts from 2007/08 onward are observed. Figure C.2 in the appendix shows the availability of data for the different age groups and birth cohorts. As there is also considerable variation in the daycare expansion speed during and after the main expansion period between the different counties, the generalized DiD approach can also be applied to the youngest age group, namely the 1–2 year olds. Thus, the instantaneous effects of the reform can be assessed. With the standard DiD approach only the longer-term effects can be evaluated.

5 Empirical results

In the following section I describe and discuss the (generalized) DiD and event study results. Further, I provide evidence for the robustness of my results and conduct heterogeneity analyses.

[30] Note, for children born before 2009, I use the county of residence observed in 2009. For all other children I use the observed county of residence when they are 1–2 years old.

5.1 DiD Results

Table 2 reports the results for the longer-term effects of the daycare expansion on children's health, obtained from estimating Equation 1. The first three panels display results for the three different sets of communicable diseases: Infections, ear diseases and respiratory diseases. The fourth and fifth panel show the results for mental disorders and obesity, respectively. Lastly, the bottom two panels represent the results for healthcare consumption measured by annual treatment cases and healthcare costs. Column 1 shows the results for 3–5 year old children, column 2 for 6–8 year olds and column 3 for 9–10 year olds.
Communicable diseases. For all three sets of communicable diseases, being affected by the expansion leads to a decrease in the number of diagnoses across all age groups. However, the effects for ear diseases are statistically not significant and quite small in magnitude. For infections, the effects are only marginally significant (10% level) for children aged 6–8 years. This effect corresponds to a 2% decrease compared to the pre-treatment mean. For respiratory diseases the effects for elementary school children (aged 6–8 and 9–10) are significant on the 1% and 10% significance level and correspond to a 1.8% and 1.3% decrease.

Mental disorders and obesity. For mental disorders and obesity, the table shows a different picture: The expansion led to an increase in the number of diagnoses for these sets of conditions across all age groups. However, the effects are statistically not significant.

Healthcare consuption and costs. For healthcare consumption, I observe a significant decrease (0.036) in the number of annual treatment cases for 3–5 year old children (0.6% compared to the pre-treatment mean). For 6–8 year old children the estimate is negative as well, while the estimate for 9–10 year old children suggests an increase in treatment cases. However, these effects are smaller in magnitude and statistically not significant. Despite the significant changes in the frequency of diagnoses and healthcare consumption, I do not depict a significant effect of the expansion on healthcare costs for any age groups.

Age-specific results. Table D.2 displays the results for all age groups (3–10 years) separately. The expansion seems to increase the number of infections and ear diseases at age three, and decrease them for all other age groups. However, the effects are statistically not significant except for the negative effects on infections at age 5 and 7. The effects on respiratory diseases are negative for all age groups and turn significant at age 5 and from age 7 onward. The effects on treatment cases are negative for children aged 4–5 and particularly large

Table 2: DiD Results

	Age: 3-5	Age: 6-8	Age: 9-10
Infections	−0.006	-0.015^{+}	−0.005
	(0.009)	(0.008)	(0.007)
Pre-Treatment Mean	1.003	0.78	0.665
Ear diseases	−0.002	−0.004	−0.001
	(0.005)	(0.004)	(0.003)
Pre-Treatment Mean	0.929	0.461	0.289
Respiratory diseases	−0.013	-0.035^{**}	-0.022^{+}
	(0.014)	(0.013)	(0.012)
Pre-Treatment Mean	2.951	1.951	1.634
Mental disorders	0.011	0.017	0.008
	(0.007)	(0.012)	(0.013)
Pre-Treatment Mean	0.881	0.959	0.975
Obesity	0.0004	0.002	0.002
	(0.001)	(0.002)	(0.002)
Pre-Treatment Mean	0.043	0.061	0.095
Treatment cases	-0.036^{*}	−0.014	0.016
	(0.018)	(0.035)	(0.026)
Pre-Treatment Mean	6.44	5.357	4.941
Healthcare costs	−0.055	0.149	1.491
	(1.085)	(1.597)	(2.163)
Pre-Treatment Mean	304.959	247.641	247.662
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses/treatment cases/costs. Costs are fee-adjusted. *Source:* KBV, own calculations.

in size and statistically significant for 5 year old children. For 6 and 7 year old children the effects are also negative, while for all other age groups the effects are positive, but statistically not significant.

First-stage results. Next, I compare the significant decreases in the number of infections, respiratory diseases and treatment cases to the first stage results (Table 3), i.e. the change in the daycare coverage rate following the expansion, to compute the TT effects. The expansion led for all age groups to a significant increase in the expansion rate (25.2% for 3–5 year olds, 39.1% for 6–8 year olds and 37% for 9–10 year olds compared to the pre-treatment mean). These estimates can be interpreted as a 25.2 (39.1 or 37) percent higher probability of being in daycare at age 1–2 years for children from treatment counties compared to control counties. Scaling the ITT effects for respiratory diseases with the first stage effects, leads to a TT effect of almost 5% for 6–8 year old and of almost 4% for 9–10 year old children. Similarly, the TT effect on infections for 6–8 year olds amounts to about

5%. Thus, daycare exposure on children that reside in the treatment area and were born in post-reform cohorts, i.e. have a 39% higher likelihood of being in daycare compared to their counterparts, decreases the number of infections and respiratory diagnoses for elementary school children by about 5%. The TT effect for treatment cases for 3–5 year old children amounts to a reduction of 2.4%.

Table 3: DiD Results: First stage

	Age: 3-5	Age: 6-8	Age: 9-10
Coverage rate	2.712***	3.025***	2.218***
	(0.155)	(0.258)	(0.315)
Effect in %	25.2%	39.1%	37.0%
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	11,504,915	15,484,013	9,711,011
ITT respiratory diseases	−0.4%	−1.8%**	−1.3%+
TT respiratory diseases	−1.6%	−4.6%**	−3.5%+
ITT infections	−0.6%	−1.9%+	−0.8%
TT infections	−2.4%	−4.9%+	−2.2%
ITT treatment cases	−0.6%*	−0.3%	0.3%
TT treatment cases	−2.4%*	−0.8%	0.8%

Note: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Generalized DiD. As a next step, I turn to the results of the generalized DiD approached obtained from estimating Equation 3 displayed in Table 4. Generally, the results show a similar picture as the DiD results, i.e. decreases of communicable diseases in the longer run and mostly insignificant or small estimates for mental disorders and obesity. Specifically, the generalized DiD results depict a positive instantaneous effect of the expansion on all three communicable diseases, i.e. the number of diagnoses increases for 1–2 year old children. The effects are all highly significant (on the 0.1% significance level) and amount to 0.009 for infections, 0.004 for ear diseases and 0.018 for respiratory diseases. This means that an increase of the daycare coverage rate by 1 percentage point increases the number of infections by 0.009. The estimates correspond to a 0.6% increase for infections, 0.7% for ear diseases and 0.6% for respiratory diseases compared to the sample mean. In line with the DiD results, the results depict negative and significant effects for respiratory diseases for elementary school children (0.3% for both 6–8 year olds and 9–10 year olds compared to the sample means). The effects for ear diseases are only marginally significant or small in size while this approach reveals negative and significant effects on infections for elementary school children (0.4% for 6–8 year olds and 0.6% for 9–10 year olds). For respiratory diseases, the increase at age 1–2 years appears to be larger than the accumulated decreases at age 6–8 and 9–10 years. For infections, the increase at age 1–2 years is equal in size compared to the decrease at elementary school age.

The estimates for mental disorders are insignificant for all age groups except for 9–10 year olds. Here, results show a decrease of 0.5% compared to the sample mean. For obesity, I obtain positive and significant increases for children below school age (0.9% for 1–2 year olds and 0.5% for 3–5 year olds) and a significant decrease for 9–10 year olds (1%). For health care consumption, the results point out an increase for 1–2 year olds (0.1%) and a decrease for all other age groups (0.2% for 3–5 year olds and 0.2% for 6–8 year olds). Here, the decreases from age 3–8 appear to be larger than the increase at age 1–2 years. The estimates for healthcare costs point in the same direction as for healthcare consumption: positive for 1–2 year olds and negative for 3–10 year olds. The effects are all highly significant, but very small in magnitude. Thus they do not provide evidence for an economically meaningful effect on healthcare costs.

Table 4: Generalized DiD Results

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Infections	0.009^{***}	0.001	-0.003^{***}	-0.004^{***}
	(0.002)	(0.001)	(0.001)	(0.001)
Sample Mean	1.394	1	0.777	0.665
Ear diseases	0.004^{***}	-0.001^{*}	-0.001^{*}	-0.001^{+}
	(0.001)	(0.001)	(0.0004)	(0.0003)
Sample Mean	0.583	0.84	0.454	0.284
Respiratory diseases	0.018^{***}	-0.001	-0.005^{***}	-0.004^{*}
	(0.003)	(0.002)	(0.001)	(0.001)
Sample Mean	2.854	2.653	1.852	1.583
Mental disorders	0.001	0.0003	-0.002	-0.005^{**}
	(0.001)	(0.001)	(0.002)	(0.002)
Sample Mean	0.312	0.867	1.057	1.031
Obesity	0.0002^{*}	0.0002^{+}	-0.0002	-0.001^{***}
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Sample Mean	0.022	0.043	0.063	0.098
Treatment cases	0.008^{*}	-0.013^{***}	-0.008^{**}	-0.005
	(0.004)	(0.003)	(0.002)	(0.003)
Sample Mean	6.331	6.135	5.282	4.915
Healthcare costs	1.097^{***}	-0.331^{*}	-0.661^{***}	-0.786^{**}
	(0.182)	(0.167)	(0.161)	(0.243)
Sample Mean	320.035	287.353	244.572	249.226
Control for age + gender	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009
Observations	8,522,334	14,117,111	13,979,440	9,455,394

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses/treatment cases/costs. Costs are fee-adjusted. *Source:* KBV, own calculations.

5.2 Event-Study Graphs

TO COME

5.3 Detailed Diagnoses

So far, I have presented results for quite broad sets of diseases. In this section, I provide results for more narrowly defined diagnoses (3- and 4-digit level of ICD-10 codes) within these broad sets of diseases. In particular, I selected diagnoses that are captured within the studied sets of diseases and belong to the 50 most frequently reported diagnoses by pediatricians (see ZI (2015)).

Infections. Within infections, intestinal infectious diseases are responsible for about 15% of infections of 6–8 year old children. Herein other gastroenteritis and colitis of infectious and unspecified origin (e.g. "abdominal influenza") account for about 80%. Intestinal infections and therein gastroenteritis are the only subgroups among the studied subgroups for which I can depict significant effects for 3–10 year old children (Table F.3 and F.6). Namely, the expansion led to a decrease of 0.004 intestinal infections and gastroenteritis diagnoses for 9–10 year old children and to a decrease of 0.005 gastroenteritis diagnoses for 6–8 year old children. The effect for 9–10 year old children is similar in size compared to the effect on all infections (despite the general effect not being significant). For 1–2 year old children I find significant increases for all studied subgroups (Table F.6).

Ear diseases. For the aggregated set of ear diseases I do not find sizable changes following the daycare expansion for 3–10 year old children. However, the more detailed analysis reveals that the reform led to a significant and sizable decrease in the number of suppurative and unspecified otitis media for 3–8 year old children which account for about one third of all ear diseases (Table F.3). For 1–2 year old children I find an increase in overall ear diseases, which can be attributed to increases in diseases of middle ear and mastoid (and herein nonsuppurative otitis media) and other disorders of the ear (Table F.6).

Respiratory diseases. The decrease in respiratory diseases for elementary school age children can be mostly attributed to diseases of the upper respiratory tract, chronic lower respiratory diseases and other acute lower respiratory infections, which account combined for almost half of all respiratory conditions for 6–8 year old children (Table F.4 and F.7). Increases in respiratory diseases at age 1–2 years are particularly prevalent for acute upper respiratory infections, other acute lower respiratory infections (herein acute bronchitis) and other diseases of upper respiratory tract (herein chronic rhinitis).

Mental disorders. The analysis of the effects of the expansion on the aggregated set of mental disorders does not reveal any significant effects. However, when looking into frequent subgroups within the standard DiD framework, I find evidence that the expansion increased the number of development disorders (of speech and language) across all age groups (Table F.5). The generalized DiD analysis only depicts significant increases for developmental disorders for 1–2 year old children (Table F.8).

Multiple hypothesis testing. Given the large number of health outcomes used in the analysis, I perform multiple-hypothesis-tests using a Bonferroni correction adjustments procedure to the single physical and mental health diagnoses. I correct for 33 hypotheses (number of detailed diagnoses considered).[31] The underlying regressions are the DiD and generalized DiD specifications from Equation 1 and 3. When applying the Bonferroni correction procedure, within the set of *infections*, only the effect for 1–2 year old children on intestinal infectious diseases and the effects for 9–10 year old children on intestinal infectious diseases and gastroenteritis diagnoses remain statistically significant. The multiple hypothesis method confirms my finding for most diagnoses within *ear diseases*, *respiratory diseases* and *mental disorders*.

5.4 Alternative specifications

Alternative treatment definitions. In my baseline specification, I define treatment and control counties by ordering counties by the percentage point increase during the main expansion period and separate the sample at the median. In order to test the sensitivity of my results towards this assumption, I run the analysis with other definitions. Specifically, I compare the upper 40% with the lower 40%, the upper 35% with the lower 35%, the upper 30% with the lower 30%, the upper 25% with the lower 25% and the upper 20% with the lower 20%. In these specifications, children from counties below the upper threshold and above the lower threshold are excluded from the analysis. Tables G.9 – G.15 display the results for the different outcomes. The coefficients for ear diseases, mental disorders and healthcare costs remain insignificant across most specifications. The results on respiratory diseases and treatment cases are hardly different from the main results across the different treatment definitions. Interestingly, the effects on infections of 6–8 year old children are

[31] I choose the Bonferroni corrrection as my preferred method since this is the most conservative correction procedure. I implement this by using the R-packge *p.adjust*.

highly statistically significant for all specifications except for my main specification (separation at the median). This supports my conclusion that the expansion did not only affect the number of respiratory diseases but also infections. The marginally significant effect in the main specification can be explained by the fact, that my main specification is the most conservative measure as this includes the comparison of children from counties that have very similar expansion rates (just below the median vs. just above the median). The results on obesity display positive effects across all specification which are mostly insignificant. This matches the uncertain results form the main analysis. Thus, I am not able to draw conclusions on the effects of the expansion on obesity.

Alternative expansion period definitions. To test the sensitivity of my results towards the exact definition of the expansion period, I run the regressions with other definitions, defining 2008–2011, 2009–2012 and 2009–2013 as the main expansion period. The results are displayed in Table G.16-G.22. The results on ear diseases, mental disorders and obesity remain largely small and statistically insignificant. For infections and respiratory diseases, the estimates for 6–8 year olds are similar in size and significance to my baseline results across all three definitions. While the results for treatment cases for the alternative expansion period 2009–2013 matches the baseline results, the results differ for the expansion period definitions 2009–2012 and 2008–2011. In both cases, the coefficient for 3–5 year olds is not significant, while there is a positive and highly significant effect for 6–8 year olds. The results for different expansion periods for healthcare costs match the baseline results, i.e. they do not depict any significant effects.

Alternative assumptions for county of residence. TO COME

Common trend assumption. In order to causally interpret my results, the common trend assumption needs to hold. I provide evidence for the plausibility of this assumption with the event study graphs presented in Section 5.2. To further prove the validity of this assumption, I exclude time and county fixed effects as well as other control variables from the specification in Equation 1. The results shown in Table G.23 are very similar to my baseline specification. Furthermore, I exclude the $phase-in$ dummy in the regressions for 6–10 year old children. Results also for respiratory diseases do not change compared to the main results (Table G.24). However, the coefficients on infections decrease in size and turn insignificant.

Furthermore, in order to provide evidence that my results reflect a reform effect and not just some underlying time trend, I conduct a placebo analysis. I choose diabetes mellitus as an outcome variable which is a chronic disease that should not be affected by environmental factors such as daycare attendance. The results show for all age groups very small coefficients which are statistically not significant (Table G.25). To make sure that the results are not driven by secular changes between urban and rural areas coinciding with the reform, I further drop all cities (Kreisfreie Städte) with more than 500,000 inhabitants. The results hardly change compared to the main results (Table G.26). Taken together, all these tests support the plausibility of the common trend assumption and thereby justify the causal interpretation of my results.

Randomization Inference. My results stay robust when I take the varying cluster sizes into account and evaluate the effects more conservatively by applying randomization inference. I discuss the results only for the outcomes for which I find evidence for a significant impact of the expansion. For completeness, all graphs can be found in the Appendix. When comparing the original estimates (red, vertical lines) with the randomization distributions (gray bars), the significance of the effects can be determined. For respiratory diseases, the graphical inspection clearly shows that it is feasible to test a one-sided alternative and compare the original estimates to more extreme values to the right (Figure G.3). For the 3–5 year olds the upper p-value amounts to 0.1 (Panel (a)), for the 6–8 year olds to 0.02 (Panel (b)) and for the 9–10 year olds to 0.05 (Panel (c)).[32]

Extensive margin. In my main specification I investigate the effect of the reform on the number of diagnoses. Alternatively, I am interested in the development of the prevalence of the diagnoses, i.e. what share of children has a relevant diagnosis at least once per year.[33] The results are presented in Table G.27 for the DiD and in Table G.28 for the generalized DiD specification. The DiD results are very similar to the main results, i.e. depicting significant decreases in the prevalence of respiratory diseases for 6–8 year old children. While the coefficients for infections are still negative, they are not statistically significant. The results of the generalized DiD displays significant increases in the prevalence of all three communicable diseases at age 1–2 and decreases for infections and respiratory diseases for older age groups. Thus, these results support my main message, i.e. that children have a

[32]Note, due to computation power, the randomization inference analyses are based on randomly drawn 1.5 million observations and only 100 repetitions. In the future I plan to run it for the whole sample with 1000 repetitions. So far I only included only the results for respiratory diseases, the results for the remaining outcomes will follow as this project proceeds.

[33]The outcomes variables are defined as dummy variables.

higher chance of having infections and respiratory diseases at age 1–2 and a lower risk at elementary school age following the daycare expansion.

5.5 Discussion of the Results

In sum, the results on communicable diseases provide some evidence for a day-care driven intertemporal substitution of illness spells, from the first years of elementary school towards the first years of daycare. More precisely, on the one hand children suffer more frequently from infections, ear diseases and respiratory diseases when entering daycare at age 1–2. On the other hand, at elementary school age children fall less often sick of these conditions. These results are intuitive as children that catch an infection, build antibiodies against these viruses or bacteria. Over time, they become immune against more and more infections and therefore fall less often sick. When entering daycare children are in close contact to other children and therefore exposed to many viruses and bacteria. Entering daycare earlier leads to worse health in the short run, but initiates the immunization process earlier leading to less infections in the longer run. For infections the increase when entering daycare is comparable to the decrease at elementary school age. In contrast, the increase in respiratory diseases at age 1–2 years is about twice as large as the decrease at elementary school age.

These results are in line with the hygiene hypothesis, van den Berg and Siflinger (2020) as well as findings from the medical literature (e.g. Enserink et al., 2013).[34] The results of van den Berg and Siflinger (2020) are similar in the sense that they also find increases in infections and respiratory diseases following daycare exposure in the short-run. However, they find more pronounced effects on ear diseases in the longer run, while my effects are mostly focused on infections and respiratory diseases. However, the three sets of conditions are closely related and different coding practices and daycare environments in Sweden and Germany could explain these differences. Furthermore, the analysis of the more narrowly defined outcomes reveals also effects of the expansion on otitis media which is a subgroup of ear diseases. The general message, i.e. early daycare exposure leads to worse physical health in the short-run and better health in the longer run, is the same.

In contrast to my results, it was found that a large-scale daycare reform in Quebec led to negative effects on health both in the short and in the long-run. The effects are mostly

[34] Note, Enserink et al. (2013) does not control for selection into daycare and can therefore not be interpreted causally. However, it provides evidence for an association between attending daycare and catching infections, which is stronger at younger ages.

driven by children that had access to daycare at very young ages (Baker et al., 2008, 2019; Kottelenberg and Lehrer, 2013). These contrasting results might be driven by differences in daycare quality: While daycare in Sweden is considered high-quality and in Germany mediocre-quality, the expansion in Quebec was rather cheap and is considered low-quality care. Findings from other studies analyzing smaller and targeted programs (e.g. Perry Preschool Program and Abecedarian Program) show positive effects on short- and long-run health outcomes (e.g. Conti et al., 2016).

To get a better understanding of the effect sizes, I compare my estimates with estimates found for other factors that influence health. One important determinant of the prevalence of infections, ear diseases and respiratory diseases is age. In my data, the average annual number of diagnoses of infections and respiratory diseases are about 8% and 12% lower, respectively, for seven year old children compared to six year old children among the pre-reform cohorts (Table D.2). Although these effects are not causal estimates of age on the prevalence of these diseases, it puts my estimates into context. I find that a 39% higher likelihood of being in daycare leads to a reduction of about 5% per daycare slot of both infections and respiratory diseases for 6–8 year old children. Other factors that are known to influence health are education and second-hand smoking. In terms of education, Kemptner et al. (2011) finds that one additional year of schooling reduces long-term illness of men by 4.1 percentage points and the probability of being overweight by 3.1 percentage points. Huebener (2018) points out intergenerational effects of education on health (behavior). In particular, he provides evidence that one additional year of maternal school reduces the probability that children will smoke by 17% and to become overweight by 21%. Several smoke-free legislations in US states were shown to reduce the number of outpatient emergency department visits of teenagers for several diseases (e.g. 12% reduction for asthma, 8% reduction for ear infections and 9% reduction for upper respiratory infections) (Hawkins et al., 2016). Based on these findings, the observed reductions of communicable diseases in elementary school aged children appear sizable. One additional year of education leads three to four times bigger effect sizes compared to an increase of 39% in the probability of being in daycare at age 1–2 years while one year in age or smoke-free legislation lead to about twice as large effect sizes.

My results provide little evidence that the daycare expansion affected mental health. In Section 5.3 I evaluate, whether certain common mental disorders are differently affected which could lead to the overall null effect. Here, I provide evidence that children affected from the reform might suffer more often from development disorders at young ages. However, for all other subgroups I do not find significant effects. My findings are in contrast to van den Berg and Siflinger (2020) who find substantial decreases in the prevalence of mental disorders for almost all age groups. These differences could arise due to the differences in the counterfactual and the timing of entering daycare: While in Sweden the reform led to a change from informal care into formal care for all age groups, the expansion in Germany caused a switch from home care into daycare only for the children below the age of 3. Children in informal daycare could benefit from a switch to formal daycare where care actors are potentially more qualified. In Germany, the majority of children from aged 3 onward were in daycare already prior to the reform. Therefore, only the timing of entering daycare changed but not the daycare environment.

I also do not find clear evidence for an effect of daycare on obesity. According to the generalized DiD results, there is slight evidence for an increase in obesity at early ages and a decrease at older ages. Lauber (2015) points out that children at the margin, i.e. children whose daycare usage is affected by regional daycare provision gain from enrollment with 30 months or earlier in terms of significantly fewer weight problems. Also D'Onise et al. (2010) find in a meta study that daycare/preschool attendance leads to a reduction in obesity.

In line with more communicable diseases at young ages and fewer communicable diseases at older ages, I provide evidence that the daycare expansion led to more healthcare consumption at age 1–2 and less healthcare consumption in the longer run. This matches the results of van den Berg and Siflinger (2020). The effects on healthcare consumption are rather small in magnitude (+0.1 % and −0.2% for 1–2 year and 3–5 year olds, respectively for the generalized DiD results.[35]. One potential explanation could be, that the effect of some parents taking their child more often to the doctor to get a sick note while others taking it less often to the doctor for time reasons cancel each other out. Another reason could be, that these considerations might not have changed substantially due to the reform. Even though mothers labor force participation has increased, the effects are quite small in size (Müller and Wrohlich, 2020). However, despite the change in healthcare consumption and frequency of diagnoses, I do not depict sizable effects of the expansion on healthcare costs. This result is quite surprising, but a potential explanation lies in the billing system of the

[35]For comparison, the increase in respiratory diseases at age 1–2 corresponds to 0.6%.

German public healthcare system: Physicians get reimbursed only once for patients that show up multiple times with the same diagnoses during one quarter. Thus, more frequent doctor visits for the same diagnoses during one quarter are not captured in healthcare costs.

My results raise the question whether the substitution of illness spells for infections and respiratory diseases from the first years of elementary school to the first years of daycare is beneficial or not. In terms of healthcare costs arising in the first ten years of life, the daycare expansion appears to be neither beneficial nor costly. However, to evaluate the welfare effects in terms of health of the reform other aspects such as severity/duration of illness spells at different ages, sickness absence at school/daycare, long-term health effects and opportunity costs for parents need to be considered. I will carefully evaluate these aspects for a "back-of-the-envelope"-calculation of the welfare effects in terms of health of the reform as this project proceeds.

5.6 Effect heterogeneity

To address the question whether different socio-economic groups were differently affected by the expansion I conduct heterogeneity analyses. Due to data limitations, I am only able to observe socio-economic characteristics on the county level. I sort the counties by the value of different socio-economic characteristics and separate the sample at the median to compare counties below and above the median values. As the share of migrants in a county, populations density and GDP per capita were shown to be determinants of the treatment status, I do not check for effect heterogeneity along these dimensions as this would likely reflect the effect of the treatment.

Education. The decreases in infections and treatment cases at age 3–5 years and 6–8 years are clearly driven by children from counties where the share of citizens with a university degree is above median (Table H.29). The pattern for respiratory diseases is less clear, but hints in the same direction. Interestingly, mental disorders seem to increase in children from counties with a below median share of citizens that hold a university degree across all age groups (Table H.30). Note, this finding should be interpreted carefully as the take up rate of early child care slots is known to be higher among families with a higher level of education.

Household income. All significant effects, namely reductions in infections, respiratory diseases and treatment cases are driven by children from counties with below median household incomes. This finding is contrast to the finding for education as education and income are positively correlated.

Gender. There is no evidence for different effects of the expansion on boys and girls (Table H.31).

6 Conclusion

This paper provides novel insights about the causal effects of a large-scale daycare expansion on a multi-dimensional and comprehensive set of health outcomes. For identification, I exploit temporal and spatial variation in the expansion speed across German counties and employ difference-in-differences and event study approaches.

Despite health being one of the most important determinants for the development of both cognitive and non-cognitive skills of children, there is so far little evidence on the effects of daycare attendance on child health. The few studies that do assess the effects of daycare on health mostly rely on survey data that usually contain subjective and broad health measure (with van den Berg and Siflinger (2020) being an exception). However, health is multi-dimensional and the effects of the daycare (expansions) on different health outcomes might, therefore, go into different directions.

My analyses are based on administrative data from German health insurance funds that include health diagnoses of all publicly insured children. I use a sample of children born between 1999 and 2015 who are observed between 2009 and 2019. The data contain all diagnoses in outpatient care during the observation period. Specifically, I identify and consider relevant diagnoses and measures within three dimensions of health outcomes: physical health (communicable diseases and obesity), mental disorders, and healthcare consumption and costs.

My empirical results provide evidence that early daycare attendance increases the prevalence of respiratory and infectious diseases at age one to two but decreases the prevalence at older ages. The reductions are most pronounced at the first years of primary school (age six to eight). In line with the hygiene hypothesis, this suggests a substitution of illness spells from the first years of elementary school to the first years of daycare. The intention to treat effect amounts to about 1–2% fewer diagnoses for respiratory and infectious diseases (depending on the age group), while the treatment-on-the-treated effect is about 5%. The increase in infections at age 1–2 years corresponds roughly to the decrease at elementary school age, while the increase in respiratory diseases at early ages is about twice as large as the decrease at older ages. The observed reductions of communicable diseases in elementary school aged children appear sizable in light of other reforms or factors that affect health. For example, one additional year of education leads three to four times bigger effect sizes compared to an increase of 39% in the probability of being in daycare at age 1—2 years while one year in age or smoke-free legislation lead to about twice as large effect sizes.

For mental health and obesity, I find no significant changes. Health care consumption increases at ages 1–2 years while it decreases at ages 3–5 years. There is no effect on healthcare consumption for older children. Despite changes in the prevalence of diagnoses and number of doctor visits, there is no effect on healthcare costs for any age group. The findings are robust to a large set of robustness checks such as different definitions of the treatment status and the expansion period and the application of randomization inference methods to obtain valid p-values.

These results raise the question whether the substitution of illness spells for infections and respiratory diseases from the first years of elementary school to the first years of daycare is beneficial or not. In terms of healthcare costs arising in the first ten years of life, the daycare expansion appears to be neither beneficial nor costly. However, to evaluate the welfare effects in terms of health of the reform other aspects such as severity/duration of illness spells at different ages, sickness absence at school/daycare, long-term health effects and opportunity costs for parents need to be considered. I will carefully evaluate these aspects for a "back-of-the-envelope"-calculation of the welfare effects in terms of health of the reform as this project proceeds.

References

Aalto, A.-M., Mörk, E., Sjögren, A. and Svaleryd, H. (2019), Does childcare improve the health of children with unemployed parents? evidence from swedish childcare access reform, Technical report, Working Paper.

Baker, M., Gruber, J. and Milligan, K. (2008), 'Universal child care, maternal labor supply, and family well-being', *Journal of political Economy* **116**(4), 709–745.

Baker, M., Gruber, J. and Milligan, K. (2019), 'The long-run impacts of a universal child care program', *American Economic Journal: Economic Policy* **11**(3), 1–26.

Baron, R., Taye, M., Besseling-van der Vaart, I., Ujčič-Voortman, J., Szajewska, H., Seidell, J. C. and Verhoeff, A. (2020), 'The relationship of prenatal and infant antibiotic exposure with childhood overweight and obesity: a systematic review', *Journal of developmental origins of health and disease* **11**(4), 335–349.

Barschkett, M., Geyer, J., Haan, P. and Hammerschmid, A. (2021), 'The effect of an increase in the retirement age on health - evidence from administrative data', *Mimeo* .

Birch, L. L. (1999), 'Development of food preferences', *Annual review of nutrition* **19**(1), 41–62.

Briggs, N., Weatherhead, J., Sastry, K. J. and Hotez, P. J. (2016), 'The hygiene hypothesis and its inconvenient truths about helminth infections', *PLoS neglected tropical diseases* **10**(9), e0004944.

Cameron, A. C. and Trivedi, P. K. (2005), *Microeconometrics: methods and applications*, Cambridge university press.

Carneiro, P., Crawford, C. and Goodman, A. (2007), 'The impact of early cognitive and non-cognitive skills on later outcomes', *Centre for Economics of Education Discussion Paper* **92**.

Carneiro, P. and Ginja, R. (2014), 'Long-term impacts of compensatory preschool on health and behavior: Evidence from head start', *American Economic Journal: Economic Policy* **6**(4), 135–73.

Conti, G., Heckman, J. J. and Pinto, R. (2016), 'The effects of two influential early childhood interventions on health and healthy behaviour', *The Economic Journal* **126**(596), F28–F65.

Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. (2018), 'Who benefits from universal child care? estimating marginal returns to early child care attendance', *Journal of Political Economy* **126**(6), 2356–2409.

Cunha, F. and Heckman, J. (2007), 'The technology of skill formation', *American Economic Review* **97**(2), 31–47.

Currie, J. (2020), 'Child health as human capital', *Health economics* **29**(4), 452–463.

Currie, J. and Stabile, M. (2006), 'Child mental health and human capital accumulation: the case of adhd', *Journal of health economics* **25**(6), 1094–1118.

Davis, K. and Christoffel, K. K. (1994), 'Obesity in preschool and school-age children: treatment early and often may be best', *Archives of pediatrics & adolescent medicine* **148**(12), 1257–1261.

de Hoog, M. L., Venekamp, R. P., van der Ent, C. K., Schilder, A., Sanders, E. A., Damoiseaux, R. A., Bogaert, D., Uiterwaal, C. S., Smit, H. A. and Bruijning-Verhagen, P. (2014), 'Impact of early daycare on healthcare resource use related to upper respiratory tract infections during childhood: prospective whistler cohort study', *BMC medicine* **12**(1), 1–8.

Destatis (2018*a*), *Statistiken der Kinder-und Jugendhilfe 2006-2013: Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderten Kindertagespfelge am 01.03.*

Destatis (2018*b*), *Statistiken der Kinder-und Jugendhilfe: Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderter Kindertagespflege am 01.03. 2018.* Wiesbaden Statistisches Bundesamt.

Destatis (2019), *Kindertagesbetreuung regional.* Accessed: 2021-08-04.
URL: *https://www.statistikportal.de/de/karten/kindertagesbetreuung-regional*

D'Onise, K., Lynch, J. W., Sawyer, M. G. and McDermott, R. A. (2010), 'Can preschool improve child health outcomes? a systematic review', *Social science & medicine* **70**(9), 1423–1440.

Enserink, R., Ypma, R., Donker, G. A., Smit, H. A. and van Pelt, W. (2013), 'Infectious disease burden related to child day care in the netherlands', *The Pediatric infectious disease journal* **32**(8), e334–e340.

Felfe, C. and Lalive, R. (2018), 'Does early child care affect children's development?', *Journal of Public Economics* **159**, 33–53.

Gerber, A. S. and Green, D. P. (2012), *Field experiments: Design, analysis, and interpretation*, WW Norton.

Gesundheitsministerium (2020), *Statistik über Versicherte, gegliedert nach Status, Alter, Wohnort und Kassenart 2004-2020*. Accessed: 2021-08-05.
URL: *https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung/mitglieder-und-versicherte.html*

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* .

Havnes, T. and Mogstad, M. (2011), 'No child left behind: Subsidized child care and children's long-run outcomes', *American Economic Journal: Economic Policy* **3**(2), 97–129.

Hawkins, S. S., Hristakeva, S., Gottlieb, M. and Baum, C. F. (2016), 'Reduction in emergency department visits for children's asthma, ear infections, and respiratory infections after the introduction of state smoke-free legislation', *Preventive medicine* **89**, 278–285.

Heckman, J. J. (2006), 'Skill formation and the economics of investing in disadvantaged children', *Science* **312**(5782), 1900–1902.

Heckman, J. J. (2007), 'The economics, technology, and neuroscience of human capability formation', *Proceedings of the national Academy of Sciences* **104**(33), 13250–13255.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. and Yavitz, A. (2010), 'The rate of return to the highscope perry preschool program', *Journal of public Economics* **94**(1-2), 114–128.

Heckman, J., Pinto, R. and Savelyev, P. (2013), 'Understanding the mechanisms through which an influential early childhood program boosted adult outcomes', *American Economic Review* **103**(6), 2052–86.

Huebener, M. (2018), 'The effects of education on health: An intergenerational perspective'.

Huebener, M., Pape, A. and Spiess, C. K. (2020), ‘Parental labour supply responses to the abolition of day care fees’, *Journal of Economic Behavior & Organization* **180**, 510–543.

Hüsken, K. (2011), *Kita vor Ort. Betreuungsatlas auf Ebene der Jugendamtsbezirke 2010.*, Deutsches Jugendinstitut.

Kamper-Jørgensen, M., Wohlfahrt, J., Simonsen, J., Grønbæk, M. and Benn, C. S. (2006), ‘Population-based study of the impact of childcare attendance on hospitalizations for acute respiratory infections’, *Pediatrics* **118**(4), 1439–1446.

Kamtsiuris, P., Bergmann, E., Rattay, P. and Schlaud, M. (2007), Inanspruchnahme medizinischer leistungen : Ergebnisse des kinder- und jugendgesundheitssurveys (kiggs), *in* ‘Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz’, Vol. 50, Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung.

KBV (2019), ‘Honorarbericht’. accessed on 13/9/21.
URL: *https://www.kbv.de/html/honorarbericht.php*

Kemptner, D., Jürges, H. and Reinhold, S. (2011), ‘Changes in compulsory schooling and the causal effect of education on health: Evidence from germany’, *Journal of health economics* **30**(2), 340–354.

Kottelenberg, M. J. and Lehrer, S. F. (2013), ‘New evidence on the impacts of access to and attending universal child-care in canada’, *Canadian Public Policy* **39**(2), 263–286.

Lauber, V. (2015), Essays in Empirical Family Economics: Investments in Children and Work-Life Balance of Parents, PhD thesis.

Mbakwa, C. A., Scheres, L., Penders, J., Mommers, M., Thijs, C. and Arts, I. C. (2016), ‘Early life antibiotic exposure and weight development in children’, *The Journal of pediatrics* **176**, 105–113.

Muehler, G. (2010), ‘Consequences of mixed provision of child care: An overview on the german market’.

Müller, K.-U. and Wrohlich, K. (2016), ‘Two steps forward—one step back? evaluating contradicting child care policies in germany’, *CESifo Economic Studies* **62**(4), 672–698.

Müller, K.-U. and Wrohlich, K. (2020), ‘Does subsidized care for toddlers increase maternal labor supply? Evidence from a large-scale expansion of early childcare', *Labour Economics* **62**, 101776.

Must, A., Spadano, J., Coakley, E. H., Field, A. E., Colditz, G. and Dietz, W. H. (1999), 'The disease burden associated with overweight and obesity', *Jama* **282**(16), 1523–1529.

Nicoletti, C., Salvanes, K. G. and Tominey, E. (2020), 'Mothers working during preschool years and child skills. does income compensate?'.

OECD (2019*a*), *Education at a Glance 2019*.
URL: *https://www.oecd-ilibrary.org/content/publication/f8d7880d-en*

OECD (2019*b*), Germany: Country Health Profile 2019, State of Health in the EU, Technical report, OECD Publishing, Paris/European Observatory on Health Systems and Policies, Brussels.

OECD (2020), *OECD family database*. Accessed: 2021-01-21.
URL: *http://www.oecd.org/social/family/database.htm*

Oikonomopoulou, K., Brinc, D., Kyriacou, K. and Diamandis, E. P. (2013), 'Infection and cancer: revaluation of the hygiene hypothesis', *Clinical Cancer Research* **19**(11), 2834–2841.

Peet, E. D., McCoy, D. C., Danaei, G., Ezzati, M., Fawzi, W., Jarvelin, M.-R., Pillas, D. and Fink, G. (2015), 'Early childhood development and schooling attainment: longitudinal evidence from british, finnish and philippine birth cohorts', *PloS one* **10**(9), e0137219.

Pfeifer, G., Reutter, M. and Strohmaier, K. (2020), 'Goodbye smokers' corner health effects of school smoking bans', *Journal of Human Resources* **55**(3), 1068–1104.

Sandner, M. (2019), 'Effects of early childhood intervention on fertility and maternal employment: Evidence from a randomized controlled trial', *Journal of health economics* **63**, 159–181.

Schaub, B., Lauener, R. and von Mutius, E. (2006), 'The many faces of the hygiene hypothesis', *Journal of allergy and clinical immunology* **117**(5), 969–977.

Schienkiewitz, A., Brettschneider, A.-K., Damerow, S. and Schaffrath Rosario, A. (2018), 'Overweight and obesity among children and adolescents in germany. results of the cross-sectional kiggs wave 2 study and trends'.

Schmitz, S., Spieß, C. K. and Stahl, J. F. (2017), 'Day care centers: Family expenditures increased significantly at some points between 1996 and 2015', *DIW Economic Bulletin* **7**(42), 411–423.

Spiess, C. K. (2008), 'Early childhood education and care in germany: The status quo and reform proposals', *Journal of Business Economics, ZfB Special* **1**, 4–27.

Spieß, C. K. (2011), 'Vereinbarkeit von familie und beruf–wie wirksam sind deutsche „care policies "?', *Perspektiven der Wirtschaftspolitik* **12**, 4–27.

Statista (2021), *Anzahl der Geburten in Deutschland von 1991 bis 2020*. Accessed: 2021-07-21.
URL: *https://de.statista.com/statistik/daten/studie/235/umfrage/anzahl-der-geburten-seit-1993/*

Strachan, D. P. (1989), 'Hay fever, hygiene, and household size.', *BMJ: British Medical Journal* **299**(6710), 1259.

van den Berg, G. J. and Siflinger, B. M. (2020), The effects of day care on health during childhood: evidence by age, IZA Discussion Paper 12891.

Watamura, S. E., Coe, C. L., Laudenslager, M. L. and Robertson, S. S. (2010), 'Child care setting affects salivary cortisol and antibody secretion in young children', *Psychoneuroendocrinology* **35**(8), 1156–1166.

Waters, E., de Silva-Sanigorski, A., Burford, B. J., Brown, T., Campbell, K. J., Gao, Y., Armstrong, R., Prosser, L. and Summerbell, C. D. (2011), 'Interventions for preventing obesity in children', *Cochrane database of systematic reviews* (12).

Wrohlich, K. (2008), 'The excess demand for subsidized child care in germany', *Applied Economics* **40**(10), 1217–1228.

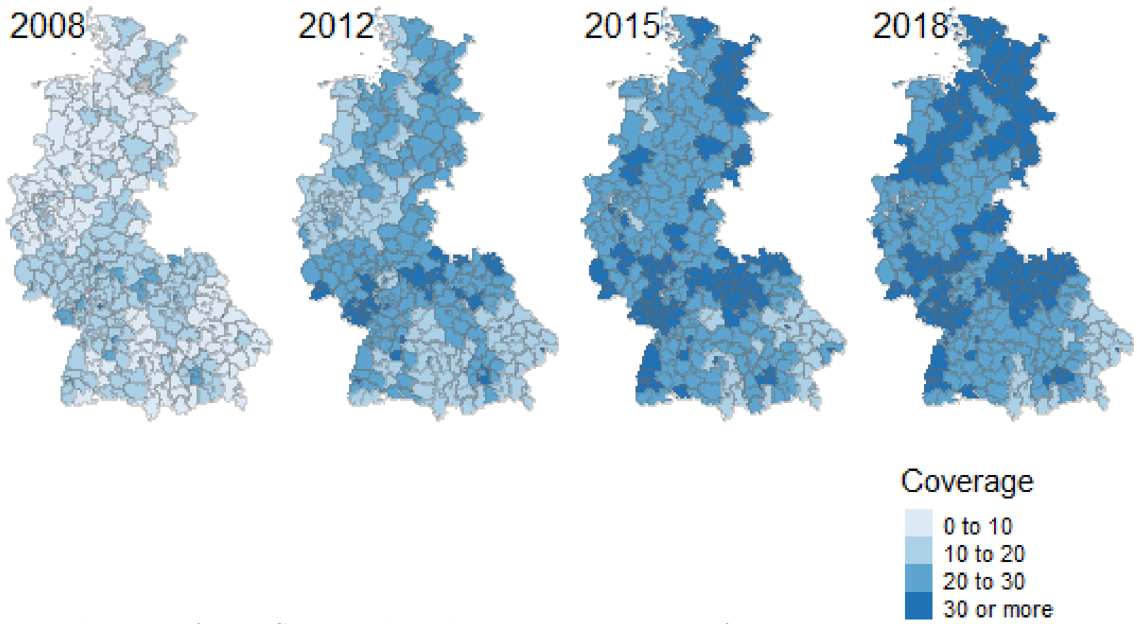
Zerpa, M. (2018), 'Short and medium run impacts of preschool education: Evidence from state pre-k programs'.

ZI (2015), Die 50 häufigsten icd-10-schlüsselnummern nach fachgruppen, Technical report, Zentralinstitut für die Kassenärztliche Versorgung in Deutschland, Ansprechpartner: Joachim Heuer.

Appendices

A Regional variation in the daycare expansion

Figure A.1: Daycare expansion during 2008 and 2018



Note: The maps of West-Germany show the daycare coverage rate for children below three years per county for different years in percent.
Source: Destatis 1994-2018, own calculations

B County characteristics

Table B.1: Descriptive statistics treatment vs. control counties

	Control counties (N = 163)	Treatment counties (N = 160)
Daycare coverage rate		
mean (sd)	10.9853 ± 4.7558	12.3356 ± 4.5809
Unemployment rate		
mean (sd)	6.3000 ± 2.9573	5.8000 ± 2.2728
Share of population U3		
mean (sd)	2.5153 ± 0.2095	2.4356 ± 0.2250
Average age		
mean (sd)	42.4270 ± 1.1592	42.5744 ± 1.2664
Share of migrants		
mean (sd)	9.9798 ± 4.5270	6.7119 ± 3.0633
Fertility rate		
mean (sd)	1.4025 ± 0.1065	1.4144 ± 0.1051
Infant mortality		
mean (sd)	3.7123 ± 2.0819	3.4644 ± 1.9653
Life expectancy		
mean (sd)	80.2098 ± 1.0357	80.1806 ± 0.8335
Female employment rate		
mean (sd)	44.9859 ± 4.0261	45.8387 ± 3.3097
Household income		
mean (sd)	1,624.4294 ± 213.8849	1,593.3500 ± 172.5046
Population density		
mean (sd)	760.8405 ± 803.7271	357.5312 ± 463.3122
GDP per capita		
mean (sd)	33.2859 ± 12.5471	28.0613 ± 12.4159
Excess nitrogen		
mean (sd)	76.1969 ± 24.2337	71.0931 ± 26.4032

Note: Means ± standard deviations are reported. *Source:* INKAR, own calculations.

C Data availability

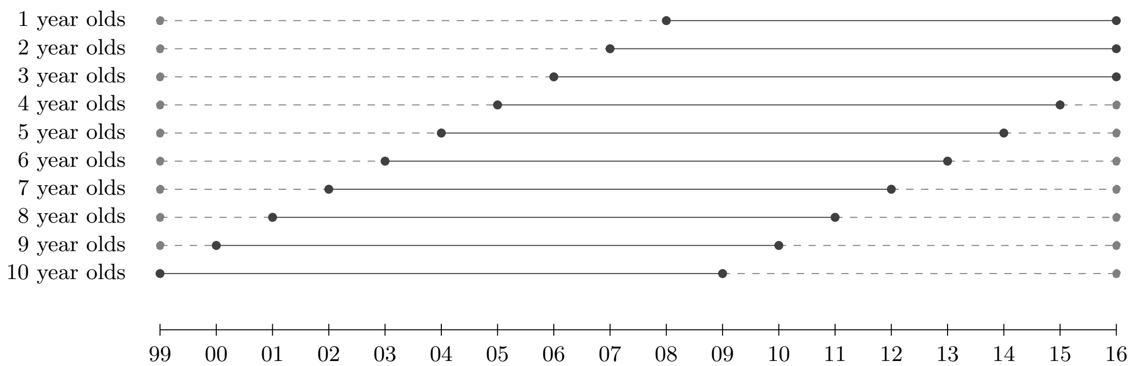


Figure C.2: Data availability by age group and birth cohort

Note: The graph shows data availability for different birth cohorts and age groups,.

Source: KBV 1999-2016, own calculations

D DiD by age group

Table D.2: DiD Results by age group

	Age: 3	Age: 4	Age: 5	Age: 6	Age: 7	Age: 8	Age: 9	Age: 10
Infections	0.008	−0.012	-0.018^{+}	−0.013	-0.017^{+}	−0.008	−0.007	−0.006
	(0.011)	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.008)	(0.007)
Pre-Treatment Mean	1.121	1.039	0.94	0.854	0.783	0.729	0.686	0.647
Ear diseases	0.009	−0.001	−0.006	−0.003	−0.004	0.001	−0.001	−0.0001
	(0.007)	(0.007)	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)
Pre-Treatment Mean	0.842	1.071	0.865	0.633	0.449	0.357	0.309	0.272
Respiratory diseases	−0.004	−0.027	-0.030^{+}	−0.026	-0.033^{*}	-0.033^{*}	-0.030^{*}	-0.026^{*}
	(0.016)	(0.018)	(0.016)	(0.017)	(0.015)	(0.014)	(0.012)	(0.011)
Pre-Treatment Mean	2.386	2.542	2.261	1.96	1.725	1.613	1.539	1.535
Mental disorders	0.014^{+}	0.014	0.010	0.016	0.013	0.016	0.014	0.006
	(0.007)	(0.009)	(0.010)	(0.013)	(0.012)	(0.014)	(0.015)	(0.014)
Pre-Treatment Mean	0.451	0.832	1.055	1.032	0.918	0.946	0.986	0.967
Obesity	−0.0003	0.001	0.001	0.0005	0.002	0.004^{*}	0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Pre-Treatment Mean	0.023	0.031	0.042	0.051	0.056	0.073	0.084	0.103
Treatment cases	0.007	-0.048^{*}	-0.082^{***}	−0.039	−0.030	0.002	0.087	0.102
	(0.025)	(0.022)	(0.019)	(0.035)	(0.029)	(0.027)	(0.027)	(0.026)
Pre-Treatment Mean	6.41	6.619	6.331	5.865	5.306	5.063	4.972	4.915
Healthcare costs	2.115	−0.090	-2.404^{+}	−0.979	−0.328	2.214	2.105	0.937
	(1.643)	(1.255)	(1.260)	(1.733)	(1.560)	(2.132)	(2.338)	(2.322)
Pre-Treatment Mean	286.316	318.977	301.784	259.63	241.473	244.825	248.977	246.524
Control for gender	yes	yes	yes	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Birth cohorts	2006-2011	2005-2011	2004-2011	2003-2011	2002-2011	2001-2011	2000-2010	1999-2009
Observations	3,085,275	3,600,165	4,128,673	4,666,085	5,200,958	5,733,931	5,806,039	5,868,828

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses/treatment cases/costs. Costs are fee-adjusted. *Source:* KBV, own calculations.

E Event-Study graphs

TO COME

F Detailed diagnoses

Table F.3: DiD Results: Detailed Diagnoses (infections and ear diseases)

	Age: 3-5	Age: 6-8	Age: 9-10
Infections			
Intestinal infectious diseases (A00-A09)	0.001	−0.004	−0.004*
	(0.003)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	1	0.840
Pre-Treatment Mean	0.157	0.113	0.099
Other gastroenteritis and colitis of infectious and unspecified origin (A09)	−0.0005	−0.005+	−0.004*
	(0.003)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	1	0.569
Pre-Treatment Mean	0.125	0.09	0.08
Other viral diseases (B25-B34)	0.0002	−0.001	−0.001
	(0.004)	(0.005)	(0.004)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.258	0.16	0.126
Viral infection of unspecified site (B34)	0.0001	−0.001	−0.001
	(0.005)	(0.005)	(0.004)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.249	0.154	0.121
Other infectious diseases (B99-B99)	−0.005	−0.003	0.002
	(0.003)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.197	0.11	0.077
Other and unspecified infectious diseases (B99)	−0.005	−0.003	0.002
	(0.003)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.197	0.11	0.077
Ear diseases			
Diseases of middle ear and mastoid (H65-H75)	−0.005	−0.003	0.002
	(0.003)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	0.400	1
Pre-Treatment Mean	0.197	0.11	0.077
Suppurative and unspecified otitis media (H66)	−0.007**	−0.003**	−0.001
	(0.002)	(0.001)	(0.001)
Bonferroni-corrected p-value	0.075	0.241	1
Pre-Treatment Mean	0.296	0.125	0.072
Nonsuppurative otitis media (H65)	0.003	−0.002	−0.001
	(0.003)	(0.002)	(0.001)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.302	0.116	0.056
Other disorders of ear (H90-H95)	−0.005	−0.003	0.002
	(0.003)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	0.400	1
Pre-Treatment Mean	0.197	0.11	0.077
Other hearing loss (H91)	0.001	0.001	0.0003
	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.043	0.023	0.016
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	11,504,915	15,484,013	9,711,011

Note: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses. *Source:* KBV, own calculations.

Table F.4: DiD Results: Detailed Diagnoses (respiratory diseases)

	Age: 3-5	Age: 6-8	Age: 9-10
Respiratory diseases			
Acute upper respiratory infections (J00-J06)	0.001	−0.002	0.001
	(0.002)	(0.003)	(0.003)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.648	0.502	0.431
Acute nasopharyngitis (common cold, J00)	0.001	−0.001	−0.001
	(0.004)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.191	0.109	0.081
Acute pharyngitis (J02)	−0.0001	−0.002	−0.001
	(0.003)	(0.002)	(0.002)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.149	0.115	0.103
Acute tonsillitis (J03)	−0.0003	0.001	0.003
	(0.003)	(0.002)	(0.002)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.267	0.18	0.13
Acute laryngitis, tracheitis (J04)	0.003	0.001	0.001
	(0.002)	(0.002)	(0.001)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.081	0.049	0.039
Acute upper respiratory infections of multiple and unspecified sites (J06)	0.004	0.001	0.004
	(0.007)	(0.006)	(0.005)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.807	0.492	0.389
Other acute lower respiratory infections (J20-J22)	−0.001	-0.006^{+}	-0.007^{*}
	(0.004)	(0.004)	(0.003)
Pre-Treatment Mean	0.385	0.206	0.146
Bonferroni-corrected p-value	1	1	1
Acute bronchitis (J20)	0.001	-0.006^{+}	-0.006^{+}
	(0.004)	(0.003)	(0.003)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.347	0.185	0.131
Other diseases of upper respiratory tract (J30-J39)	-0.008^{+}	-0.010^{*}	-0.008^{+}
	(0.005)	(0.005)	(0.004)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.559	0.39	0.365
Vasomotor and allergic rhinitis (J30)	−0.001	−0.004	−0.005
	(0.002)	(0.003)	(0.003)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.083	0.154	0.214
Chronic rhinitis, nasopharyngitis and pharyngitis (J31)	-0.007^{*}	−0.002	−0.001
	(0.003)	(0.002)	(0.002)
Bonferroni-corrected p-value	0.984	1	1
Pre-Treatment Mean	0.132	0.077	0.057
Chronic lower respiratory diseases (J40-J47)	-0.012^{*}	-0.022^{**}	-0.017^{**}
	(0.005)	(0.007)	(0.006)
Bonferroni-corrected p-value	0.858	0.051	0.282
Pre-Treatment Mean	0.437	0.359	0.337
Bronchitis, not specified as acute or chronic (J40)	-0.007^{+}	−0.006	−0.001
	(0.004)	(0.005)	(0.005)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.205	0.114	0.081
Other diseases of the respiratory system (J95-J99)	0.001	0.002	0.002
	(0.004)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.178	0.098	0.071
Other respiratory disorders (J98)	0.001	0.002	0.002
	(0.004)	(0.003)	(0.002)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.176	0.097	0.07
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	11,504,915	15,484,013	9,711,011

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses. *Source:* KBV, own calculations.

Table F.5: DiD Results: Detailed Diagnoses (mental disorders)

	Age: 3-5	Age: 6-8	Age: 9-10
Mental disorders			
Disorders of psychological development (F80-F89)	0.010^{+}	0.009	0.003
	(0.006)	(0.007)	(0.007)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.633	0.51	0.361
Specific developmental disorders of speech and language (F80)	0.010^{*}	0.006	−0.002
	(0.005)	(0.006)	(0.005)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.516	0.345	0.179
Specific developmental disorder of motor function (F82)	−0.001	0.004	−0.0003
	(0.002)	(0.003)	(0.003)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.103	0.105	0.072
Mixed specific developmental disorders (F83)	0.002	0.007^{*}	0.006^{*}
	(0.002)	(0.003)	(0.003)
Bonferroni-corrected p-value	1	0.612	0.428
Pre-Treatment Mean	0.045	0.055	0.044
Unspecified disorder of psychological development (F89)	−0.001	−0.0005	0.0003
	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.034	0.03	0.022
Behavioural and emotional disorders with onset usually occurring	0.002	0.005	0.001
in childhood and adolescence (F90-F98)	(0.002)	(0.004)	(0.006)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.147	0.304	0.423
Other behavioural and emotional disorders with onset	0.0003	−0.0001	−0.001
usually occurring in childhood and adolescence (F98)	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	1	1	1
Pre-Treatment Mean	0.041	0.054	0.054
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	11,504,915	15,484,013	9,711,011

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses. *Source:* KBV, own calculations.

Table F.6: Generalized DiD Results: Detailed Diagnoses (infections and ear diseases)

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Infections				
Intestinal infectious diseases (A00-A09)	0.002^{***}	0.001^{*}	-0.001^{**}	-0.001^{**}
	(0.001)	(0.0004)	(0.0002)	(0.0002)
Bonferroni-corrected p-value	0.01	0.562	0.124	0.043
Pre-Treatment Mean	0.27	0.184	0.122	0.103
Other gastroenteritis and colitis of infectious and unspecified origin (A09)	0.002^{**}	0.0005	-0.001^{**}	-0.001^{***}
	(0.001)	(0.0004)	(0.0002)	(0.0002)
Bonferroni-corrected p-value	0.137	1	0.153	0.006
Pre-Treatment Mean	0.218	0.15	0.101	0.085
Other viral diseases (B25-B34)	0.002^{+}	-0.001	-0.001	-0.001^{+}
	(0.001)	(0.001)	(0.0004)	(0.0004)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.358	0.265	0.164	0.128
Viral infection of unspecified site (B34)	0.002^{+}	-0.001	-0.001	-0.001^{+}
	(0.001)	(0.001)	(0.0004)	(0.0004)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.35	0.257	0.158	0.123
Other infectious diseases (B99-B99)	0.002^{*}	0.001	-0.0003	-0.0003
	(0.001)	(0.0004)	(0.0003)	(0.0002)
Bonferroni-corrected p-value	0.904	1	1	1
Pre-Treatment Mean	0.269	0.181	0.106	0.075
Other and unspecified infectious diseases (B99)	0.002^{*}	0.001	-0.0003	-0.0003
	(0.001)	(0.0004)	(0.0003)	(0.0002)
Bonferroni-corrected p-value	0.904	1	1	1
Pre-Treatment Mean	0.269	0.181	0.106	0.075
Ear diseases				
Diseases of middle ear and mastoid (H65-H75)	0.002^{*}	0.001	-0.0003	-0.0003
	(0.001)	(0.0004)	(0.0003)	(0.0002)
Bonferroni-corrected p-value	0.001	0.014	0.001	0.895
Pre-Treatment Mean	0.469	0.587	0.275	0.152
Suppurative and unspecified otitis media (H66)	0.001	-0.001^{**}	-0.0001	0.00002
	(0.0005)	(0.0004)	(0.0002)	(0.0001)
Bonferroni-corrected p-value	1	0.037	1	1
Pre-Treatment Mean	0.266	0.252	0.115	0.068
Nonsuppurative otitis media (H65)	0.002^{***}	-0.001^{*}	-0.001^{***}	-0.0002^{+}
	(0.0005)	(0.0004)	(0.0002)	(0.0001)
Bonferroni-corrected p-value	0.001	0.705	0.002	1
Pre-Treatment Mean	0.192	0.286	0.116	0.055
Other disorders of ear (H90-H95)	0.002^{*}	0.001	-0.0003	-0.0003
	(0.001)	(0.0004)	(0.0003)	(0.0002)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.048	0.133	0.088	0.062
Other hearing loss (H91)	0.00001	0.00005	-0.00003	-0.0001
	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.009	0.036	0.023	0.016
Control for age + gender	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009
Observations	8,522,334	14,117,111	13,979,440	9,455,394

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses. *Source:* KBV, own calculations.

Table F.7: Generalized DiD Results: Detailed Diagnoses (respiratory diseases)

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Respiratory diseases				
Acute upper respiratory infections (J00-J06)	0.004**	−0.001	0.001	−0.0001
	(0.001)	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	0.075	1	1	1
Pre-Treatment Mean	1.417	1.221	0.824	0.664
Acute nasopharyngitis (common cold, J00)	0.001	0.001	0.0004	−0.0002
	(0.001)	(0.001)	(0.0004)	(0.0003)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.253	0.179	0.108	0.082
Acute pharyngitis (J02)	−0.0002	−0.00005	0.0002	−0.00004
	(0.001)	(0.0004)	(0.0003)	(0.0003)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.138	0.135	0.109	0.099
Acute tonsillitis (J03)	0.0002	−0.0001	0.001*	0.0005*
	(0.0005)	(0.0004)	(0.0003)	(0.0002)
Bonferroni-corrected p-value	1	1	1	0.851
Pre-Treatment Mean	0.149	0.218	0.165	0.123
Acute laryngitis, tracheitis (J04)	0.0002	0.001	−0.00003	−0.0002
	(0.0004)	(0.0004)	(0.0002)	(0.0001)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.077	0.067	0.043	0.036
Acute upper respiratory infections of	0.004**	−0.001	−0.0003	−0.0004
multiple and unspecified sites (J06)	(0.001)	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	0.209	1	1	1
Pre-Treatment Mean	1.018	0.795	0.493	0.39
Other acute lower respiratory infections (J20-J22)	0.005***	0.001	−0.001*	−0.001*
	(0.001)	(0.001)	(0.0005)	(0.0004)
Bonferroni-corrected p-value	0.000	1	1	0.846
Pre-Treatment Mean	0.466	0.362	0.199	0.144
Acute bronchitis (J20)	0.004***	0.0003	−0.001*	−0.001*
	(0.001)	(0.001)	(0.0005)	(0.0004)
Bonferroni-corrected p-value	0.001	1	1	0.802
Pre-Treatment Mean	0.419	0.328	0.179	0.129
Other diseases of upper respiratory tract (J30-J39)	0.003**	−0.001+	−0.002***	−0.002***
	(0.001)	(0.001)	(0.0005)	(0.001)
Bonferroni-corrected p-value	0.085	1	0.001	0.004
Pre-Treatment Mean	0.352	0.489	0.378	0.361
Vasomotor and allergic rhinitis (J30)	0.0004	−0.0003	−0.001**	−0.002***
	(0.0003)	(0.0002)	(0.0003)	(0.0004)
Bonferroni-corrected p-value	1	1	0.034	0.001
Pre-Treatment Mean	0.021	0.067	0.151	0.215
Chronic rhinitis, nasopharyngitis and pharyngitis (J31)	0.002*	0.0003	0.0001	0.0001
	(0.001)	(0.001)	(0.0003)	(0.0002)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.188	0.118	0.074	0.055
Chronic lower respiratory diseases (J40-J47)	0.002	−0.0001	−0.003***	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	1	1	0.017	1
Pre-Treatment Mean	0.319	0.34	0.312	0.309
Bronchitis, not specified as acute or chronic (J40)	−0.001	0.001	0.0001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.157	0.117	0.07	0.057
Other diseases of the respiratory system (J95-J99)	0.002+	0.001+	0.0004	−0.0001
	(0.001)	(0.001)	(0.0003)	(0.0003)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.23	0.168	0.095	0.069
Other respiratory disorders (J98)	0.002+	0.001	0.0004	−0.00003
	(0.001)	(0.001)	(0.0003)	(0.0003)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.228	0.166	0.094	0.068
Control for age + gender	yes	yes	yes	
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009
Observations	8,522,334	14,117,111	13,979,440	9,455,394

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses. *Source:* KBV, own calculations.

Table F.8: Generalized DiD Results: Detailed Diagnoses (mental disorders)

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Mental disorders				
Disorders of psychological development (F80-F89)	0.003***	0.001	−0.001	−0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	0.000	1	1	1
Pre-Treatment Mean	0.193	0.613	0.584	0.393
Specific developmental disorders of speech and language (F80)	0.002***	0.001+	−0.0004	−0.001
	(0.0003)	(0.001)	(0.001)	(0.001)
Bonferroni-corrected p-value	0.000	1	1	1
Pre-Treatment Mean	0.078	0.501	0.413	0.202
Specific developmental disorder of motor function (F82)	0.001*	−0.0005	−0.0004	0.0001
	(0.0004)	(0.0004)	(0.0004)	(0.0003)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.079	0.103	0.121	0.079
Mixed specific developmental disorders (F83)	0.0003*	0.001+	0.0001	0.00003
	(0.0001)	(0.0003)	(0.0004)	(0.0003)
Bonferroni-corrected p-value	0.796	1	1	1
Pre-Treatment Mean	0.014	0.05	0.07	0.052
Unspecified disorder of psychological development (F89)	0.0002	−0.00001	0.0003	−0.00002
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Bonferroni-corrected p-value	1	1	1	1
Pre-Treatment Mean	0.035	0.063	0.072	0.052
Behavioural and emotional disorders with onset usually occurring	−0.001	−0.0005	−0.001*	−0.003***
in childhood and adolescence (F90-F98)	(0.001)	(0.0005)	(0.001)	(0.001)
Bonferroni-corrected p-value	1	1	0.555	0.004
Pre-Treatment Mean	0.05	0.153	0.319	0.439
Other behavioural and emotional disorders with onset	−0.001**	−0.0001	−0.001**	−0.001**
usually occurring in childhood and adolescence (F98)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
Bonferroni-corrected p-value	0.233	1	0.186	0.037
Pre-Treatment Mean	0.019	0.061	0.112	0.119
Control for age + gender	yes	yes	yes	
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009
Observations	8,522,334	14,117,111	13,979,440	9,455,394

Note: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. All variables are count variables including the annual number of diagnoses. *Source:* KBV, own calculations.

G Robustness

Different Treatment definitions

Table G.9: DiD Results Infections: Different treatment definitions

	Age: 3-5	Age: 6-8	Age: 9-10
upper 40 vs. lower 40%	−0.014	-0.022^{*}	−0.009
	(0.009)	(0.009)	(0.008)
Observations	7,326,287	11,064,983	8,379,657
upper 35 vs. lower 35%	-0.018^{+}	-0.026^{**}	−0.011
	(0.010)	(0.010)	(0.009)
Observations	6,216,208	9,401,154	7,135,816
upper 30 vs. lower 30%	-0.020^{+}	-0.027^{**}	−0.008
	(0.011)	(0.009)	(0.009)
Observations	5,397,090	8,150,249	6,174,301
upper 25 vs. lower 25%	-0.025^{*}	-0.033^{**}	−0.012
	(0.012)	(0.011)	(0.010)
Observations	4,560,843	6,877,191	5,198,154
upper 20 vs. lower 20%	−0.017	-0.026^{*}	−0.008
	(0.013)	(0.012)	(0.011)
Observations	3,577,402	5,378,709	4,057,549
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.10: DiD Results Ear diseases: Different treatment definitions

	Age: 3-5	Age: 6-8	Age: 9-10
upper 40 vs. lower 40%	−0.001	−0.004	0.0002
	(0.006)	(0.004)	(0.003)
Observations	7,326,287	11,064,983	8,379,657
upper 35 vs. lower 35%	−0.003	−0.005	−0.001
	(0.006)	(0.004)	(0.003)
Observations	6,216,208	9,401,154	7,135,816
upper 30 vs. lower 30%	0.001	−0.003	0.002
	(0.007)	(0.004)	(0.003)
Observations	5,397,090	8,150,249	6,174,301
upper 25 vs. lower 25%	−0.002	−0.004	0.003
	(0.008)	(0.005)	(0.004)
Observations	4,560,843	6,877,191	5,198,154
upper 20 vs. lower 20%	0.001	−0.004	0.005
	(0.009)	(0.005)	(0.004)
Observations	3,577,402	5,378,709	4,057,549
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.11: DiD Results Respiratory diseases: Different treatment definitions

	Age: 3-5	Age: 6-8	Age: 9-10
upper 40 vs. lower 40%	-0.026	-0.048^{**}	-0.028^{*}
	(0.016)	(0.015)	(0.013)
Observations	7,326,287	11,064,983	8,379,657
upper 35 vs. lower 35%	-0.029^{+}	-0.041^{*}	-0.025^{+}
	(0.017)	(0.016)	(0.015)
Observations	6,216,208	9,401,154	7,135,816
upper 30 vs. lower 30%	-0.017	-0.034^{+}	-0.017
	(0.019)	(0.018)	(0.016)
Observations	5,397,090	8,150,249	6,174,301
upper 25 vs. lower 25%	-0.025	-0.046^{*}	-0.027
	(0.021)	(0.020)	(0.018)
Observations	4,560,843	6,877,191	5,198,154
upper 20 vs. lower 20%	-0.025	-0.055^{*}	-0.031^{+}
	(0.024)	(0.022)	(0.019)
Observations	3,577,402	5,378,709	4,057,549
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.12: DiD Results Mental disorders: Different treatment definitions

	Age: 3-5	Age: 6-8	Age: 9-10
upper 40 vs. lower 40%	0.013^{+}	0.020	0.012
	(0.008)	(0.013)	(0.015)
Observations	7,326,287	11,064,983	8,379,657
upper 35 vs. lower 35%	0.014^{+}	0.021	0.012
	(0.008)	(0.014)	(0.016)
Observations	6,216,208	9,401,154	7,135,816
upper 30 vs. lower 30%	0.019^{*}	0.023	0.014
	(0.009)	(0.015)	(0.017)
Observations	5,397,090	8,150,249	6,174,301
upper 25 vs. lower 25%	0.020^{*}	0.021	0.016
	(0.010)	(0.017)	(0.018)
Observations	4,560,843	6,877,191	5,198,154
upper 20 vs. lower 20%	0.018	0.023	0.028
	(0.011)	(0.019)	(0.020)
Observations	3,577,402	5,378,709	4,057,549
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.13: DiD Results Obesity: Different treatment definitions

	Age: 3-5	Age: 6-8	Age: 9-10
upper 40 vs. lower 40%	0.0003	0.002	0.002
	(0.001)	(0.002)	(0.002)
Observations	7,326,287	11,064,983	8,379,657
upper 35 vs. lower 35%	0.001	0.002	0.001
	(0.001)	(0.002)	(0.002)
Observations	6,216,208	9,401,154	7,135,816
upper 30 vs. lower 30%	0.001	0.003^{+}	0.003
	(0.001)	(0.002)	(0.002)
Observations	5,397,090	8,150,249	6,174,301
upper 25 vs. lower 25%	0.001	0.004^{*}	0.004
	(0.002)	(0.002)	(0.003)
Observations	4,560,843	6,877,191	5,198,154
upper 20 vs. lower 20%	0.003	0.007^{***}	0.008^{**}
	(0.002)	(0.002)	(0.003)
Observations	3,577,402	5,378,709	4,057,549
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.14: DiD Results Treatment cases: Different treatment definitions

	Age: 3-5	Age: 6-8	Age: 9-10
upper 40 vs. lower 40%	-0.065^{***}	-0.033	0.013
	(0.019)	(0.042)	(0.030)
Observations	7,326,287	11,064,983	8,379,657
upper 35 vs. lower 35%	-0.067^{**}	-0.015	0.026
	(0.021)	(0.050)	(0.035)
Observations	6,216,208	9,401,154	7,135,816
upper 30 vs. lower 30%	-0.064^{**}	-0.002	0.040
	(0.023)	(0.057)	(0.038)
Observations	5,397,090	8,150,249	6,174,301
upper 25 vs. lower 25%	-0.077^{**}	-0.006	0.038
	(0.025)	(0.066)	(0.043)
Observations	4,560,843	6,877,191	5,198,154
upper 20 vs. lower 20%	-0.072^{*}	0.002	0.035
	(0.028)	(0.082)	(0.051)
Observations	3,577,402	5,378,709	4,057,549
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.15: DiD Results healthcare costs: Different treatment definitions

	Age: 3-5	Age: 6-8	Age: 9-10
upper 40 vs. lower 40%	−1.054	0.133	2.306
	(1.231)	(1.865)	(2.455)
Observations	7,326,287	11,064,983	8,379,657
upper 35 vs. lower 35%	−1.206	0.174	3.213
	(1.261)	(1.896)	(2.682)
Observations	6,216,208	9,401,154	7,135,816
upper 30 vs. lower 30%	−0.872	1.052	5.139^{+}
	(1.347)	(1.966)	(2.702)
Observations	5,397,090	8,150,249	6,174,301
upper 25 vs. lower 25%	−1.182	0.977	5.812^{*}
	(1.509)	(2.170)	(2.845)
Observations	4,560,843	6,877,191	5,198,154
upper 20 vs. lower 20	−1.018	2.155	7.468^{*}
	(1.648)	(2.639)	(3.427)
Observations	3,577,402	5,378,709	4,057,549
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Different Expansion period definitions

Table G.16: DiD Results Infections: Different expansion period definitions

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	−0.003	-0.014^{+}	−0.007
	(0.008)	(0.008)	(0.008)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	-0.018^{*}	-0.020^{*}	-0.013^{+}
	(0.009)	(0.008)	(0.007)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640
Exp. period: 2009–2013	−0.011	-0.016^{+}	−0.009
	(0.010)	(0.008)	(0.008)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	10,817,715	13,979,440	10,605,640
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.17: DiD Results Ear diseases: Different expansion period definitions

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	0.008	0.004	0.002
	(0.005)	(0.004)	(0.003)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	−0.004	−0.002	0.0003
	(0.005)	(0.004)	(0.003)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640
Exp. period: 2009–2013	−0.002	−0.002	−0.002
	(0.005)	(0.004)	(0.003)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	10,817,715	13,979,440	10,605,640
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.18: DiD Results Respiratory diseases: Different expansion period definitions

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	−0.003	-0.025^{+}	−0.012
	(0.013)	(0.013)	(0.012)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	−0.019	-0.034^{*}	-0.020^{+}
	(0.013)	(0.013)	(0.012)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640
Exp. period: 2009–2013	−0.022	-0.037^{**}	−0.011
	(0.014)	(0.013)	(0.012)
Observations	10,817,715	13,979,440	10,605,640
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.19: DiD Results Mental disorders: Different expansion period definitions

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	0.010	0.023*	0.012
	(0.007)	(0.012)	(0.013)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	0.005	0.003	−0.0004
	(0.007)	(0.012)	(0.013)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640
Exp. period: 2009–2013	0.011	0.001	−0.001
	(0.008)	(0.012)	(0.014)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	10,817,715	13,979,440	10,605,640
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.20: DiD Results Obesity: Different expansion period definitions

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	0.001	0.002	0.002
	(0.001)	(0.002)	(0.002)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	−0.0002	0.001	0.001
	(0.001)	(0.001)	(0.002)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640
Exp. period: 2009–2013	0.001	0.003*	0.003
	(0.001)	(0.001)	(0.002)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	10,817,715	13,979,440	10,605,640
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.21: DiD Results Treatment cases: Different expansion period definitions

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	0.003	0.051*	−0.068***
	(0.012)	(0.020)	(0.020)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	−0.008	0.089***	−0.018
	(0.013)	(0.018)	(0.020)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640
Exp. period: 2009–2013	−0.067***	−0.050	−0.021
	(0.018)	(0.031)	(0.027)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	10,817,715	13,979,440	10,605,640
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.22: DiD Results healthcare costs: Different expansion period definitions

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	−0.069	0.319	1.517
	(1.024)	(1.565)	(2.137)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	−0.731	−2.111	−1.264
	(0.975)	(1.550)	(2.147)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640
Exp. period: 2009–2013	−1.317	−1.513	−0.879
	(1.018)	(1.576)	(2.206)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	10,817,715	13,979,440	10,605,640
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Exclusion of controls

Table G.23: DiD Results: Without controls

	Age: 3-5	Age: 6-8	Age: 9-10
Infections	-0.005	-0.015^{+}	-0.005
	(0.009)	(0.008)	(0.007)
Pre-Treatment Mean	1.003	0.78	0.665
Ear diseases	-0.002	-0.004	-0.001
	(0.005)	(0.004)	(0.003)
Pre-Treatment Mean	0.929	0.461	0.289
Respiratory diseases	-0.010	-0.035^{*}	-0.024^{*}
	(0.014)	(0.014)	(0.012)
Pre-Treatment Mean	2.951	1.951	1.634
Mental disorders	0.010	0.012	-0.0003
	(0.007)	(0.012)	(0.013)
Pre-Treatment Mean	0.881	0.959	0.975
Obesity	0.001	0.002	0.002
	(0.001)	(0.002)	(0.002)
Pre-Treatment Mean	0.043	0.061	0.095
Treatment cases	-0.035^{+}	-0.018	0.008
	(0.018)	(0.035)	(0.026)
Pre-Treatment Mean	6.44	5.357	4.941
Healthcare costs	-0.157	-0.527	0.523
	(1.109)	(1.575)	(2.078)
Pre-Treatment Mean	304.959	247.641	247.662
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Exclusion of phase-in dummy

Table G.24: DiD Results: Without Phase-in dummy

	Age: 3-5	Age: 6-8	Age: 9-10
Infections	−0.006	−0.007	−0.004
	(0.009)	(0.007)	(0.006)
Pre-Treatment Mean	1.003	0.78	0.665
Ear diseases	−0.002	−0.004	−0.0002
	(0.005)	(0.003)	(0.002)
Pre-Treatment Mean	0.929	0.461	0.289
Respiratory diseases	−0.013	−0.027*	−0.016
	(0.014)	(0.011)	(0.010)
Pre-Treatment Mean	2.951	1.951	1.634
Mental disorders	0.011	0.008	0.005
	(0.007)	(0.010)	(0.011)
Pre-Treatment Mean	0.881	0.959	0.975
Obesity	0.0004	0.002	0.002
	(0.001)	(0.001)	(0.002)
Pre-Treatment Mean	0.043	0.061	0.095
Treatment cases	−0.036*	−0.019	0.010
	(0.018)	(0.023)	(0.021)
Pre-Treatment Mean	6.44	5.357	4.941
Healthcare costs	−0.055	−0.725	0.360
	(1.085)	(1.264)	(1.842)
Pre-Treatment Mean	304.959	247.641	247.662
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Placebo Regression: Diabetes

Table G.25: Placebo Regression (DiD): Diabetes

	Age: 3-5	Age: 6-8	Age: 9-10
DiD	−0.00003	−0.0003	−0.001
	(0.0004)	(0.001)	(0.001)
Pre-Treatment Mean	0.005	0.009	0.014
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	11,504,915	15,484,013	9,711,011
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Exclusion of cities with more than 500,000 inhabitants

Table G.26: DiD Results: Exclusion of big cities

	Age: 3-5	Age: 6-8	Age: 9-10
Infections	−0.005	−0.020*	−0.010
	(0.009)	(0.008)	(0.007)
Pre-Treatment Mean	0.996	0.773	0.658
Ear diseases	−0.004	−0.005	−0.002
	(0.006)	(0.004)	(0.003)
Pre-Treatment Mean	0.938	0.461	0.288
Respiratory diseases	−0.018	−0.034*	−0.021
	(0.015)	(0.015)	(0.013)
Pre-Treatment Mean	2.98	1.953	1.631
Mental disorders	0.011	0.015	0.002
	(0.008)	(0.013)	(0.015)
Pre-Treatment Mean	0.888	0.957	0.97
Obesity	0.0004	0.002	0.002
	(0.001)	(0.002)	(0.002)
Pre-Treatment Mean	0.043	0.061	0.095
Treatment cases	−0.035+	−0.010	0.011
	(0.020)	(0.041)	(0.030)
Pre-Treatment Mean	6.466	5.357	4.932
Healthcare costs	0.128	−0.588	0.190
	(1.157)	(1.680)	(2.355)
Pre-Treatment Mean	305.332	244.761	242.928
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	7,870,341	12,020,491	9,197,151

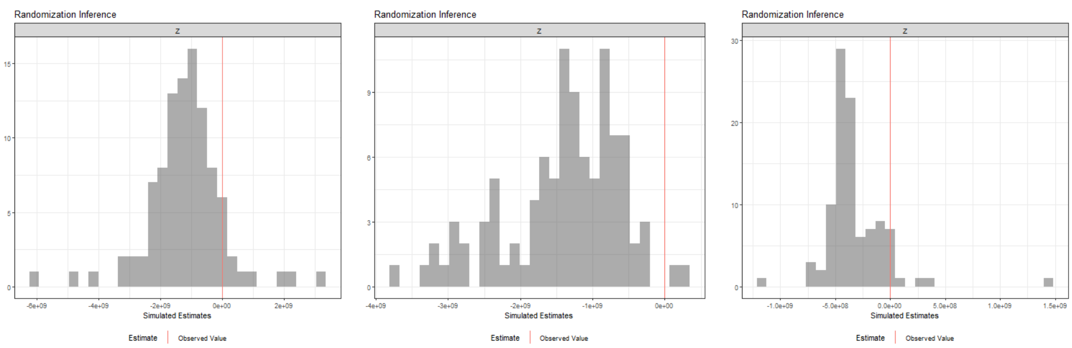
Note: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Randomization inference

Figure G.3: Randomization inference: Respiratory diseases

(a) Age: 3-5 years (b) Age: 6-8 years (c) Age: 9-10 years

Note:



The graphs show randomization inference estimates for 3-5 year old children (panel (a)), 6-8 year old children (panel (b)) and 9-10 year old children (panel (c)) for respiratory diseases. The red vertical line represents the original estimates. The grey bars display the randomization distributions. Note, due to computing power the results are based on randomly drawn 1.5 million observations and the procedure is repeated 100 times. *Source:* KBV 1999-2015, own calculations

Extensive margin

Table G.27: DiD Results: Extensive margin

	Age: 3-5	Age: 6-8	Age: 9-10
Infections	−0.002	−0.005	−0.001
	(0.003)	(0.003)	(0.003)
Pre-Treatment Mean	0.536	0.457	0.404
Ear diseases	−0.001	−0.001	−0.0003
	(0.002)	(0.001)	(0.001)
Pre-Treatment Mean	0.424	0.244	0.168
Respiratory diseases	−0.002	-0.006^{*}	−0.003
	(0.002)	(0.002)	(0.003)
Pre-Treatment Mean	0.798	0.663	0.594
Mental disorders	0.002	−0.001	−0.001
	(0.002)	(0.002)	(0.002)
Pre-Treatment Mean	0.38	0.304	0.263
Obesity	0.0003	0.001^{*}	0.002^{+}
	(0.0005)	(0.001)	(0.001)
Pre-Treatment Mean	0.025	0.032	0.05
Control for age + gender	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table G.28: Generalized DiD Results: Extensive margin

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Infections	0.003^{***}	0.0003	-0.001^{**}	-0.001^{***}
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Pre-Treatment Mean	0.63	0.534	0.456	0.404
Ear diseases	0.002^{***}	−0.0003	-0.0003^{+}	−0.0002
	(0.0004)	(0.0002)	(0.0001)	(0.0001)
Pre-Treatment Mean	0.327	0.394	0.239	0.164
Respiratory diseases	0.002^{***}	-0.001^{*}	-0.001^{**}	-0.001^{*}
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Pre-Treatment Mean	0.81	0.772	0.648	0.585
Mental diseases	0.001	−0.00004	-0.001^{**}	-0.001^{***}
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Pre-Treatment Mean	0.177	0.37	0.329	0.275
Obesity	0.0001^{+}	0.0001^{+}	0.00000	-0.0003^{**}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Pre-Treatment Mean	0.014	0.024	0.033	0.051
Control for age + gender	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009
Observations	8,522,334	14,117,111	13,979,440	9,455,394

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

H Heterogeneity

Table H.29: DiD Results by share of university degree in household on county level

	Above median share of university degree			Below median share of university degree		
	Age: 3-5	Age: 6-8	Age: 9-10	Age: 3-5	Age: 6-8	Age: 9-10
Infections	-0.029^{*}	-0.026^{+}	-0.013	0.014	-0.007	0.002
	(0.013)	(0.013)	(0.011)	(0.012)	(0.010)	(0.009)
Pre-Treatment Mean	1.014	0.794	0.677	0.992	0.767	0.653
Ear diseases	-0.006	-0.008	-0.004	0.002	0.0005	0.002
	(0.008)	(0.005)	(0.004)	(0.007)	(0.005)	(0.004)
Pre-Treatment Mean	0.904	0.453	0.286	0.952	0.467	0.291
Respiratory diseases	-0.009	-0.037^{+}	-0.029	-0.016	-0.027	-0.009
	(0.019)	(0.019)	(0.018)	(0.019)	(0.019)	(0.016)
Pre-Treatment Mean	2.853	1.905	1.601	3.045	1.994	1.666
Mental disorders	-0.002	-0.001	-0.013	0.024^{*}	0.034^{*}	0.031^{+}
	(0.010)	(0.018)	(0.020)	(0.011)	(0.017)	(0.017)
Pre-Treatment Mean	0.861	0.963	0.989	0.899	0.955	0.96
Obesity	0.001	0.002	0.001	0.0002	0.001	0.001
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
Pre-Treatment Mean	0.045	0.065	0.101	0.042	0.058	0.089
Treatment cases	-0.069^{**}	-0.016	-0.015	-0.004	-0.010	0.050
	(0.023)	(0.061)	(0.041)	(0.026)	(0.032)	(0.031)
Pre-Treatment Mean	6.324	5.302	4.934	6.545	5.402	4.944
Control for age + gender	yes	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009	2006-2011	2003-2011	2000-2009
Observations	4,553,282	6,738,141	5,221,058	4,646,631	7,179,242	5,336,256

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table H.30: DiD Results by household income on county level

	Above median household income			Below median household income		
	Age: 3-5	Age: 6-8	Age: 9-10	Age: 3-5	Age: 6-8	Age: 9-10
Infections	0.002	−0.002	0.009	−0.018	-0.031^{**}	-0.022^{*}
	(0.012)	(0.012)	(0.011)	(0.013)	(0.011)	(0.009)
Pre-Treatment Mean	0.983	0.764	0.647	1.022	0.796	0.682
Ear diseases	−0.006	-0.008^{+}	−0.005	0.003	0.002	0.004
	(0.008)	(0.005)	(0.004)	(0.007)	(0.006)	(0.004)
Pre-Treatment Mean	0.897	0.44	0.276	0.959	0.48	0.301
Respiratory diseases	−0.003	−0.019	−0.012	−0.024	-0.047^{*}	-0.031^{+}
	(0.019)	(0.016)	(0.016)	(0.020)	(0.021)	(0.018)
Pre-Treatment Mean	2.802	1.835	1.537	3.098	2.068	1.729
Mental disorders	0.003	−0.001	−0.009	0.016	0.027	0.018
	(0.010)	(0.016)	(0.019)	(0.011)	(0.017)	(0.018)
Pre-Treatment Mean	0.854	0.914	0.933	0.907	1.003	1.014
Obesity	0.001	0.002	0.0005	−0.001	0.001	0.001
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
Pre-Treatment Mean	0.039	0.055	0.086	0.048	0.068	0.104
Treatment cases	−0.018	0.018	0.017	-0.050^{*}	−0.039	0.022
	(0.025)	(0.060)	(0.040)	(0.025)	(0.030)	(0.032)
Pre-Treatment Mean	6.244	5.19	4.81	6.627	5.518	5.066
Control for age + gender	yes	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009	2006-2011	2003-2011	2000-2009
Observations	4,561,258	6,950,703	5,246,622	4,638,655	6,966,680	5,310,692

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.

Table H.31: DiD Results by gender

	Boys			Girls		
	Age: 3-5	Age: 6-8	Age: 9-10	Age: 3-5	Age: 6-8	Age: 9-10
Infections	−0.005	−0.017^{+}	−0.007	−0.007	−0.014	−0.003
	(0.010)	(0.009)	(0.008)	(0.009)	(0.009)	(0.008)
Pre-Treatment Mean	0.997	0.741	0.615	1.017	0.824	0.717
Ear diseases	0.00002	−0.004	−0.001	−0.005	−0.004	−0.001
	(0.007)	(0.004)	(0.003)	(0.006)	(0.004)	(0.003)
Pre-Treatment Mean	0.96	0.465	0.285	0.905	0.458	0.293
Respiratory diseases	−0.016	−0.039*	−0.023^{+}	−0.012	−0.033**	−0.022^{+}
	(0.016)	(0.016)	(0.014)	(0.016)	(0.013)	(0.012)
Pre-Treatment Mean	3.12	2.073	1.744	2.801	1.834	1.523
Mental disorders	0.018^{+}	0.024	0.009	0.005	0.008	0.006
	(0.009)	(0.015)	(0.017)	(0.007)	(0.010)	(0.012)
Pre-Treatment Mean	1.069	1.242	1.267	0.696	0.674	0.68
Obesity	0.001	0.002	0.003	0.0001	0.001	0.0005
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Pre-Treatment Mean	0.038	0.057	0.094	0.049	0.066	0.097
Treatment cases	−0.029	−0.046^{+}	0.001	−0.044*	0.015	0.031
	(0.021)	(0.025)	(0.024)	(0.020)	(0.062)	(0.036)
Pre-Treatment Mean	6.747	5.598	5.088	6.168	5.127	4.798
Control for age + gender	yes	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009	2006-2011	2003-2011	2000-2009
Observations	4,658,656	7,060,328	5,353,402	4,543,161	6,882,818	5,223,160

Note: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level are in parentheses. *Source:* KBV, own calculations.