(Unintended) consequences of social and family policies on health and well-being: Five Essays in Health and Family Economics

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades
einer Doktorin der Wirtschaftswissenschaft
doctor rerum politicarum
(Dr. rer. pol.)

am Fachbereich Wirtschaftswissenschaft
der Freien Universität Berlin

vorgelegt von
Mara Barschkett (M.Sc.)
geboren in Bergisch Gladbach

Berlin, 2022

Erstgutachterin: Prof. Dr. C. Katharina Spieß
Bundesinstitut für Bevölkerungsforschung und bis September 2021 Freie Universität Berlin, seit Oktober 2021 Johannes Gutenberg-Universität Mainz

Zweitgutachter: Prof. Dr. Peter Haan
Freie Universität Berlin und DIW Berlin

Disputation: 12.05.2023

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my first and second supervisors, C. Katharina Spieß and Peter Haan, for your continuous support, guidance, encouragement, and invaluable feedback. I very much appreciate that you provided all the freedom I needed to develop my research, but, at the same time, you closely followed my progress and had an open door any day and night time. You have largely contributed not only to my academic work but also to my professional and personal development with your critical feedback, patience, and kindness.

I cooperated with several co-authors (Johannes Geyer, Peter Haan, Anna Hammerschmid, Laura Schmitz, Sophia Schmitz, C. Katharina Spieß, and Elena Ziege) during my PhD, which taught me once more how much more fun, productive and rewarding teamwork is compared to working individually. I learned a lot from our collaboration, both from an academic and a personal perspective. Thank you for all the fruitful discussions and hard work; together, we produced papers that I am proud of.

I started my PhD journey in 2018 as part of the GC cohort at DIW Berlin. Studying in the big office with such a nice group made the courses during the first year a lot more enjoyable. Sporty lunch breaks, dinners, and other activities contributed to a great experience in terms of non-academic aspects of the PhD journey. In 2019, I started writing my dissertation, first in the Education and Family and later in the Public Economics Department. I want to thank Katharina and Peter for creating a friendly and welcoming atmosphere in the departments that always gave me the feeling that my input is appreciated and taken seriously. Furthermore, I am grateful for having had amazing colleagues who made me enjoy (almost) every day at DIW. Next to my co-authors, I would like to single out some colleagues who have mostly not (yet!) been co-authors. A special thanks goes to Clara Schäper and Elena Ziege for being the best "critical friends" I could ask for. Your constant support and 24-hour availability are invaluable assets to me. I am grateful to not only call you colleagues but also friends. I would also like to thank Jan Marcus for being my econometrics consultant and for your always kind, supportive, and positive words. Further, I am grateful to Jonas Jessen, Ludovica Gambaro, and Katharina Wrohlich for your academic guidance, feedback, and always welcoming and warm manner.

Most of my time at DIW, I shared my office with Laura Schmitz, who I luckily not only call co-author and office mate but also a friend. I enjoyed the friendly and supportive atmosphere and I am grateful for having been on this journey with you. In the last couple of months, I shared my office with Annekatrin Schrenker. Thanks for the great atmosphere and the mutual support during these last challenging months.

My PhD was partly funded through a scholarship by the Födernetzwerk Alterssicherung. I also received funding from the Ministry for Education and Research through the Bremen Initiative to Foster Early Childhood Development (BRISE, Förderkennzeichen 01NV1601B) and from the Stiftung Ravensburger Verlag. The financial support is much appreciated. On top, I am very grateful for the support of the Joachim Herz Stiftung through an Add-on fellowship. The fellowship not only funded many conference visits and parts of my research stay at the WHO in Geneva but also offers an excellent network, including other young and inspiring researchers. Furthermore, through a Carlo-Schmid scholarship, my research stay in Geneva was also financially supported by the DAAD and the Studienstiftung des Deutschen Volkes. Moreover, I am grateful for the National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesverinigung) for data access and excellent support, especially by Michael Radel, Christian Gallowitz and Diana Kurch-Bek. Lastly, I would like to acknowledge the (financial) support and encouragement provided by the DIW Graduate Center – a special thank you goes to Juliane Metzner for keeping our backs free and organizing great GC summer workshops.

Last but not least, I want to say a big thank you to my family and friends who have constantly supported me; I couldn't have done it without you. While I cannot list everyone, I particularly would like to thank Hedda Nielsen and Mats Kröger for continuing our successful support group that originated in 2016 in Stockholm. Furthermore, I am deeply grateful to my parents (Claudia and Dirk) and my sister (Nora) for your enduring support and encouragement throughout my life. Your backing, regardless of my academic success, is so valuable to me. Finally, above all, thank you, Tobi, for your unconditional support, for being my mood equalizer in challenging times, and for always encouraging me to strive for more.

Berlin, May 16, 2023
Mara Barschkett

Contents

Erklärung zu Co-Autorenschaften

Diese Dissertation besteht aus fünf (Arbeits-)Papieren, von denen vier in Zusammenarbeit mit Co-Autor*innen entstanden sind:

- Mara Barschkett, Johannes Geyer, Peter Haan und Anna Hammerschmid: ***"The Effects of an Increase in the Retirement Age on Health – Evidence from Administrative Data"***
- Mara Barschkett, C. Katharina Spieß und Elena Ziege: ***"Does Grandparenting Pay off for the Next Generations? Intergenerational Effects of Grandparental Care"***
- Mara Barschkett: ***"Age-specific Effects of Early Daycare on Children's Health"***
- Mara Barschkett, Johannes Geyer, Peter Haan und Anna Hammerschmid: ***"The Effects of an Increase in the Retirement Age on Health Care Costs – Evidence from Administrative Data"***
- Mara Barschkett, Laura Schmitz und Sophia Schmitz: ***"Costs and short-term effects of a home-visiting program in BRISE – first steps for a cost-effectiveness analysis"***

Liste der Vorveröffentlichungen

- **Publikationen in referierten Fachzeitschriften**
  - **Vorpublikation von Kapitel 2:**
    - Barschkett, M., Geyer, J., Haan, P. and Hammerschmid, A. (2022). The Effects of an Increase in the Retirement Age on Health – Evidence from Administrative Data. *The Journal of the Economics of Ageing, 23,* 100403.
      `https://doi.org/10.1016/j.jeoa.2022.100403`
  - **Vorpublikation von Kapitel 5:**
    - Barschkett, M., Geyer, J., Haan, P. and Hammerschmid, A. (2022). The Effects of an Increase in the Retirement Age on Health Care Costs – Evidence from Administrative Data. *The European Journal of Health Economics,* 1-20.
      `https://doi.org/10.1007/s10198-022-01535-w`
- **Beiträge zur Politikberatung**
  - **Basierend auf Kapitel 2 und 5 ist erschienen:**
    - Barschkett, M., Geyer, J., Haan, P. (2022). Anhebung des Renteneintrittsalters hat negative Konsequenzen für die Gesundheit. *DIW Wochenbericht, 89(41)*, 527-533.
  - **Teilweise basierend auf Kapitel 3 ist erschienen:**
    - Barschkett, M., Gambaro, L., Schäper, C., Spiess, C. K. und Ziege, E. (2022). Oma und Opa gefragt? Veränderungen in der Enkelbetreuung - Wohlbefinden von Eltern - Wohlergehen von Kindern. *BiB Bevölkerungstudien, 1/2022.*
- **Arbeitspapiere**
  - **Vorpublikationen von Kapitel 2:**

  * Barschkett, M., Geyer, J., Haan, P. and Hammerschmid, A. (2021). The Effects of an Increase in the Retirement Age on Health – Evidence from Administrative Data. *IZA DP*, 14893.
  * Barschkett, M., Geyer, J., Haan, P. and Hammerschmid, A. (2021). The Effects of an Increase in the Retirement Age on Health – Evidence from Administrative Data. *DIW Discussion Papers*, 1985.

- **Vorpublikationen von Kapitel 3:**
  * Barschkett, M., Spieß, C. K. and Ziege, E. (2021). Does Grandparenting Pay off for the Next Generations? Intergenerational Effects of Grandparental Care. *IZA DP*, 14795.
  * Barschkett, M., Spieß, C. K. and Ziege, E. (2021). Does Grandparenting Pay off for the Next Generations? Intergenerational Effects of Grandparental Care. *DIW Discussion Papers*, 1152.
  * Barschkett, M., Spieß, C. K. and Ziege, E. (2022). Does Grandparenting Pay off for the Next Generations? Intergenerational Effects of Grandparental Care. *BiB Working Paper*, 2/2022.

- **Vorpublikation von Kapitel 4:**
  * Barschkett, M. (2022). Age-specific Effects of Early Daycare on Children's Health. *DIW Discussion Papers*, 2028.
  * Barschkett, M. (2022). Age-specific Effects of Early Daycare on Children's Health. *BSE Discussion Papers*, No. 5.

Rechtliche Erklärung

Erklärung gem. §4 Abs. 2 (Promotionsordnung)

Hiermit erkläre ich, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur Überprüfung vorgelegen hat.

Berlin, 16. Mai 2023
Mara Barschkett

Erklärung gem. §10 Abs. 3 (Promotionsordnung)

Hiermit erkläre ich, dass ich für die Dissertation folgende Hilfsmittel und Hilfen verwendet habe. Auf dieser Grundlage und in Zusammenarbeit mit meinen Co-Autor*innen habe ich die Arbeit selbstständig verfasst.

- Software:
  - Stata Versionen 15, 16 und 17
  - Rstudio Versionen 1.2.1335 und 2021.09.0
  - LaTeX mit Overleaf
- Literatur: siehe Literaturverzeichnis

Berlin, 16. Mai 2023
Mara Barschkett

Abstract

This dissertation consists of five independent chapters contributing to the literature of health and family economics. The main topic concerns how social and family policies impact health and well-being of different generations – children, parents and grandparents.[1]

Chapter 2 analyzes the causal effect of an increase in the retirement age on official health diagnoses. We exploit a sizable cohort-specific pension reform for women using a difference-in-differences approach. The analysis is based on official records covering all individuals insured by the public health system in Germany and including all certified diagnoses by practitioners. This enables us to gain a detailed understanding of the multi-dimensionality in these health effects. The empirical findings reflect the multi-dimensionality but allow for deriving two broader conclusions. We provide evidence that the increase in the retirement age negatively affects health outcomes as the prevalence of several diagnoses, e.g., mental health, musculoskeletal diseases, and obesity, increases. In contrast, we do not find support for an improvement in health related to a prolonged working life. These findings are robust to sensitivity checks, and do not change when correcting for multiple hypothesis testing.

Chapter 3 takes an intergenerational perspective on the effects of an informal childcare setting – grandparental care – on parental and child outcomes. Grandparents act as the third largest caregiver after parental care and daycare in Germany, as in many Western societies. Adopting a double-generation perspective, we investigate the causal impact of this care mode on children's health, socio-emotional behavior, and school outcomes, as well as parental well-being. Based on representative German panel data sets, and exploiting arguably exogenous variations in geographical distance to grandparents, we analyze age-specific effects, taking into account alternative care modes. Our results suggest mainly null and in few cases negative effects on children's outcomes. If children

[1]Throughout this dissertation, the term "grandparents" is used for older individuals, i.e., individuals from the grandparents generation. **Chapter 3** specifically deals with grandparents, i.e., individuals that have grandchildren, while **Chapters 2 and 5** look more generally at older individuals from age 59.

three years and older are in full-time daycare or school and, in addition, regularly cared for by grandparents, they have more health and socio-emotional problems, in particular conduct problems. In contrast, our results point to positive effects on parental satisfaction with the childcare situation and mothers' satisfaction with leisure.

Chapter 4 aims to improve the understanding of the effects of daycare on children's health. 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 child development and later life success, the effect of early daycare attendance on health has received little attention in the economic literature. In this study, I investigate the impact of a large daycare expansion in Germany on children's age-specific mental and physical health outcomes. Based on a unique set of administrative health records covering 90% of the German population over a period of ten years, I exploit exogenous variation in daycare attendance induced by the expansion. My results provide evidence for the substitution of illness spells from the first years of elementary school to the first years of daycare. Specifically, I find that early daycare attendance increases the prevalence of respiratory and infectious diseases and healthcare consumption when entering daycare (1–2 years) by 5–6 percent. At elementary school age (6–10 years), the prevalence decreases by similar magnitudes. I do not find evidence for an effect of daycare attendance on mental disorders, obesity, injuries, vision problems, or healthcare costs. Heterogeneity analysis indicates more pronounced effects for children from disadvantaged areas, earlier detection of vision problems, and a reduction in obesity in these children.

Chapter 5 takes a cost perspective on the pension reform analyzed in Chapter 2. We use unique health record data that cover outpatient care and the associated costs to quantify the health care costs of a sizable increase in the retirement age in Germany. For the identification we exploit a sizable cohort-specific pension reform which abolished an early retirement program for all women born after 1951. Our results show that health care costs significantly increase by about 2.9% in the age group directly affected by the increase in the retirement age (women aged 60-62). We further show that the cost increase is mainly driven by the following specialist groups: Ophthalmologists, general practitioners (GPs), neurology, orthopedics, and radiology. While the effects are significant and meaningful on the individual level, we show that the increase in health care costs is modest relative to the positive fiscal effects of the pension reform. Specifically, we estimate an aggregate increase in the health costs of about 7.7 million euro for women born in 1952 aged 60-62 which amounts to less than 2% of the overall positive fiscal effects of the pension reform.

Chapter 6 concerns another childcare mode, namely parental care, by analyzing the costs and short-term effects of a home-visiting program on child and maternal outcomes. Home-visiting programs targeting families during pregnancy or shortly after birth can be a powerful tool to promote child and family well-being, in particular for disadvantaged families. However, little evidence exists on the (cost-)effectiveness of these programs in the European context. In this study, we present novel evidence on the costs and effects of *Pro Kind*, a home-visiting program under the Bremen Initiative to Foster Early Childhood Development (*BRISE*). *BRISE* randomly assigns an information and access treatment on the neighborhood level that nudges families in the treatment group to participate in *Pro Kind*. We exploit this random variation in an instrumental variables (IV) framework combined with entropy balancing to estimate the causal effects of the intervention on several mother and child outcomes during the first seven months of the children's lives. In addition, we provide cost estimates based on self-collected cost data. At this early stage of the intervention and due to data limitations, we cannot deduce meaningful causal effects of *Pro Kind* on child and maternal outcomes. The cost analysis suggests that *Pro Kind* is less costly than most comparable early childhood programs. Our analysis builds the basis for future cost-effectiveness and cost-benefit studies.

Zusammenfassung

Diese Dissertation besteht aus fünf unabhängigen Kapiteln, die einen Beitrag zur Literatur im Bereich der Gesundheits- und Familienökonomie leisten. Das Hauptthema ist die Frage, wie sich sozial- und familienpolitische Maßnahmen auf die Gesundheit und das Wohlbefinden von verschiedenen Generationen - Kindern, Eltern und Großeltern - auswirken.[2]

Kapitel 2 analysiert den kausalen Effekt einer Anhebung des Renteneintrittsalters auf ärztlich diagnostizierte Krankheiten. Unter Verwendung eines Differenz-in-Differenzen-Ansatzes untersuchen wir eine kohortenspezifische Rentenreform für Frauen, die das Renteneintrittsalter um drei Jahre angehoben hat. Die Analyse basiert auf administrativen Gesundheitsdaten, die alle gesetzlich Versicherten in Deutschland abdecken und alle von ambulant tätigen Ärzt*innen gestellte Diagnosen enthalten. Diese detaillierten Daten ermöglichen es uns, ein umfassendes Bild von der Multidimensionalität der Gesundheitseffekte zu erhalten. Die empirischen Ergebnisse spiegeln die Multidimensionalität wider, erlauben aber auch die Ableitung von zwei allgemeinen Schlussfolgerungen. Wir liefern Evidenz dafür, dass die Erhöhung des Renteneintrittsalters sich negativ auf die Gesundheit auswirkt, da die Prävalenz mehrerer Diagnosen, z. B. psychische Erkrankungen, Erkrankungen des Muskel-Skelett-Systems und Adipositas, zunimmt. Im Gegensatz dazu finden wir keine Belege für eine Verbesserung der Gesundheit im Zusammenhang mit einem längerem Arbeitsleben. Diese Ergebnisse sind robust gegenüber Sensitivitätsprüfungen und ändern sich auch nicht, wenn die p-Werte für die große Anzahl getesteter Hypothesen korrigiert werden.

Kapitel 3 befasst sich aus einer intergenerationalen Perspektive mit den Auswirkungen einer informellen Kinderbetreuungsform - der Großelternbetreuung - auf elterliche und kindliche Ergebnisvariablen. Großeltern sind in Deutschland, wie in vielen westlichen Ländern, die drittwichtigste Betreuungsform nach den Eltern und Kitas.

[2]In dieser Dissertation wird der Begriff "Großeltern"durchgängig für ältere Personen, d. h. Personen aus der Großelterngeneration, verwendet. **Kapitel 3** befasst sich speziell mit Großeltern, d. h. mit Personen, die Enkelkinder haben, während sich **Kapitel 2 und 5** mit älteren Personen ab 59 Jahren im Allgemeinen befassen.

Ausgehend von einer Zwei-Generationen-Perspektive untersuchen wir die kausalen Auswirkungen dieser Betreuungsform auf die Gesundheit, das sozio-emotionale Verhalten und die schulischen Leistungen der Kinder sowie auf das Wohlbefinden der Eltern. Auf der Grundlage repräsentativer deutscher Paneldatensätze und unter Ausnutzung exogener Variationen in der geografischen Entfernung zu den Großeltern analysieren wir altersspezifische Effekte unter Berücksichtigung alternativer Betreuungsformen. Unsere Ergebnisse deuten hauptsächlich auf Nulleffekte und in wenigen Fällen auf negative Auswirkungen auf die Entwicklung von Kindern hin. Wenn Kinder, die drei Jahre und älter sind, ganztätig eine Kita oder Schule besuchen und zusätzlich regelmäßig von den Großeltern betreut werden, weisen sie mehr gesundheitliche und sozio-emotionale Probleme auf, insbesondere Verhaltensprobleme. Im Gegensatz dazu deuten unsere Ergebnisse auf positive Auswirkungen auf die Zufriedenheit der Eltern mit der Kinderbetreuungssituation und die Zufriedenheit der Mütter mit ihrer Freizeitgestaltung hin.

Kapitel 4 zielt darauf ab, das Verständnis für die Auswirkungen von Kindertagesstätten (Kitas) auf die Gesundheit von Kindern zu verbessern. In den letzten Jahrzehnten ist der Anteil der Kleinkinder in Kitas in vielen OECD-Ländern, auch in Deutschland, deutlich gestiegen. Trotz der Bedeutung von Kindergesundheit für die kindliche Entwicklung und den späteren Erfolg im Leben wurden die Auswirkungen des Besuchs einer Kita auf die Gesundheit in der ökonomischen Literatur bisher wenig beachtet. In dieser Studie untersuche ich die Auswirkungen eines massiven Ausbaus der Kindertagesbetreuung in Deutschland auf die altersspezifische psychische und physische Gesundheit von Kindern. Auf der Grundlage einzigartiger administrativer Gesundheitsdaten, die 90% der deutschen Bevölkerung über einen Zeitraum von zehn Jahren abdecken, nutze ich durch den Kita-Ausbau ausgelöste exogene Variation in der Kita-Nutzung. Meine Ergebnisse liefern Belege für eine Substitution von infektiösen Erkrankungen von den ersten Grundschuljahren zu den ersten Jahren in der Kita. Insbesondere stelle ich fest, dass der frühe Besuch einer Kita die Prävalenz von Atemwegs- und Infektionskrankheiten sowie die Inanspruchnahme von Gesundheitsleistungen bei Eintritt in die Kita (1 bis 2 Jahre) um 5 bis 6 Prozent erhöht. Im Grundschulalter (6–10 Jahre) sinkt die Prävalenz um ähnliche Größenordnungen. Ich finde keine Belege für Auswirkungen eines Kita-Besuchs auf psychische Störungen, Adipositas, Verletzungen, Sehprobleme oder Gesundheitskosten. Eine Heterogenitätsanalyse zeigt, dass die Auswirkungen bei Kindern aus benachteiligten Gebieten ausgeprägter sind, Sehprobleme früher erkannt werden und die Fettleibigkeit bei diesen Kindern abnimmt.

In **Kapitel 5** wird die in Kapitel 2 analysierte Rentenreform aus einer Kostenperspektive betrachtet. Wir verwenden einzigartige Gesundheitsdaten, die die ambulante

Versorgung und die damit verbundenen Kosten abdecken, um die Gesundheitskosten einer beträchtlichen Anhebung des Renteneintrittsalters in Deutschland zu quantifizieren. Zur Identifizierung nutzen wir eine umfangreiche kohortenspezifische Rentenreform, die ein Frühverrentungsprogramm für alle nach 1951 geborenen Frauen abschaffte. Unsere Ergebnisse zeigen, dass die Gesundheitskosten in der von der Erhöhung des Renteneintrittsalters direkt betroffenen Altersgruppe (Frauen im Alter von 60-62 Jahren) signifikant um etwa 2,9% steigen. Wir zeigen ferner, dass der Kostenanstieg hauptsächlich von den folgenden Facharztgruppen verursacht wird: Augenärzt*innen, Allgemeinmediziner*innen, Neurologie, Orthopädie und Radiologie. Obwohl die Auswirkungen auf individueller Ebene signifikant und bedeutsam sind, zeigen wir, dass der Anstieg der Gesundheitskosten im Vergleich zu den positiven fiskalischen Auswirkungen der Rentenreform moderat ist. Konkret schätzen wir einen Gesamtanstieg der Gesundheitskosten von etwa 7,7 Millionen Euro für 1952 geborene Frauen im Alter von 60-62 Jahren, was weniger als 2% der gesamten positiven fiskalischen Auswirkungen der Rentenreform ausmacht.

Kapitel 6 befasst sich mit einer anderen Form der Kinderbetreuung, nämlich der elterlichen Betreuung, indem die Kosten und kurzfristigen Auswirkungen eines Hausbesuchsprogramms auf die Ergebnisse von Kindern und Müttern analysiert werden. Hausbesuchsprogramme, die sich an Familien während der Schwangerschaft oder kurz nach der Geburt richten, können ein wirksames Instrument zur Förderung des Wohlergehens von Kindern und Familien sein, insbesondere für benachteiligte Familien. Es gibt jedoch nur wenige Hinweise auf die (Kosten-) Effektivität dieser Programme im europäischen Kontext. In dieser Studie präsentieren wir neue Erkenntnisse über die Kosten und Auswirkungen von *Pro Kind*, einem Hausbesuchsprogramm im Rahmen der Bremer Initiative zur Förderung der frühkindlichen Entwicklung (*BRISE*). Im Rahmen von *BRISE* werden zufällig (randomisiert auf Stadtteilebene) einige Familien über *Pro Kind* informiert und ihnen wird der Zugang zu diesem Programm erleichtert. Somit werden diese Familienangeregt an *Pro Kind* teilzunehmen. Wir nutzen diese zufällige Variation in einem Instrumentalvariablen (IV)-Ansatz in Kombination mit entropy balancing, um die kausalen Effekte der Intervention auf verschiedene Ergebnisse von Müttern und Kindern während der ersten sieben Lebensmonate der Kinder zu schätzen. Darüber hinaus liefern wir Kostenschätzungen auf der Grundlage selbst erhobener Kostendaten. In diesem frühen Stadium der Intervention und aufgrund von Datenbeschränkungen können wir keine aussagekräftigen kausalen Effekte von *Pro Kind* auf die Ergebnisse bei Kindern und Müttern ableiten. Die Kostenanalyse legt nahe, dass *Pro Kind* weniger kostspielig ist als die meisten vergleichbaren frühkindlichen Programme. Unsere Analyse bildet die Grundlage für künftige Kosten-Effektivitäts- und Kosten-Nutzen-Studien.

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

Health and well-being are not only human rights[1] but also important drivers of human capital accumulation (e.g., Eide and Showalter, 2011), productivity (e.g., Bubonya et al., 2017; Stewart et al., 2003), labor force participation (e.g., Frijters et al., 2014; Pelkowski and Berger, 2004), and, ultimately, economic growth (e.g., Bloom et al., 2004; Well, 2007). Simultaneously, many western societies invest significant shares of their budget directly in health: The OECD average of health expenditures accounts for almost 10% of countries' GDP (OECD, 2022b). However, the total societal costs of ill health are likely much higher once we consider indirect costs such as productivity losses. To get an understanding of these indirect costs, it is essential to acknowledge the connection between health and economic indicators. This connection runs both ways, as health not only affects various key economic indicators but also vice-versa. Therefore, policies concerning one domain, for example labor force participation, potentially also influence health, an often overlooked consequence. In this light, how to promote health and well-being concerns policymakers and researchers alike, as they need to optimize both health and non-health indicators under the budget constraints set by the social welfare system.

Health and well-being are two closely related concepts that are difficult to distinguish from each other. In 1948, the World Health Organization (WHO) defined health as "a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity" (WHO, 1948), thereby taking a broader perspective than traditional medical definitions, defining health as "the absence of any disease or impair-

[1] Health and well-being are also adressed in the third United Nations Sustainable Development Goal that sets the aim to "ensure healthy lives and promote well-being for all at all ages" (UN, 2022).

ment" (Sartorius, 2006).[2] To this day, the WHO maintains this definition. Throughout this introduction, I adopt the WHO definition and consider well-being as an integral and essential component of health. In the Chapters following the introduction, I explicitly state how health and well-being are measured, thereby indicating which aspect of health is in the focus.

Within the economic literature on health, the original microeconomic model by Grossman saw health as a stock variable that varies over the life course in response to investment and depreciation (Grossman, 1972). One key feature of the model is that the effects of investments in early childhood fade over time as health itself depreciates. While such a health investment perspective remains central within economics, later models corrected Grossman's prediction of fading returns of early investment. Heckman and colleagues have been at the forefront of this change of perspective. Their models allow for differential and long-lasting effects of investments made at different points in time, suggesting that health today begets future health (e.g., Cunha and Heckman, 2007; Currie and Almond, 2011). Empirical evidence confirms that early health investments repercuss over the life course, i.e., good health in childhood also leads to better adult (health) outcomes (see, e.g., Currie and Almond, 2011, for an overview).[3]

Although investments in health and well-being during early childhood are particularly relevant, investments need to be sustained over the life course. Especially investing in the health of older people pays off both in the form of individual and societal benefits from good health and well-being[4], but also from a cost perspective. While those aged between 15 and 65 would, on average, incur health costs of 2,789 Euros per year, this number triples for the group aged 65 to 85 years (Destatis, 2021), suggesting large financial gains from improving the health of older people, especially in light of an aging society. These numbers capture only the costs covered by the healthcare system, however, bad health and well-being likely also translate into costs to other parts of the social welfare system, such as the pension system through disability or invalidity benefits. Societal benefits can also be examined in terms of intergenerational spill-over effects. For example, healthier and happier parents have healthier children and impact child development positively (e.g., Berger and Spiess, 2011; Coneus and Spiess, 2012a; Dahlen,

[2] The WHO definition of health and well-being has been subject to criticism (see, e.g., Godlee, 2011; Huber et al., 2011) and as of today, there exists a plethora of definitions developed by institutions and researchers of various fields.

[3] Also the epidemiological literature highlights the relevance of early life health for health during later life stages (see, e.g., Kuh et al., 2003, for an overview).

[4] There is a large body of research showing that health and well-being impact labor market outcomes such as employment, wages, and productivity (e.g., Frijters et al., 2014; Pelkowski and Berger, 2004; Stewart et al., 2003).

2016; Kelly and Bartley, 2010). Grandparents (and, more generally, older adults) with better health are more likely to provide (high quality) grandparental childcare (e.g., Del Boca et al., 2005; Zamberletti et al., 2018) and are less in need of receiving care from partners or children; thus freeing up the resources of these primary caregivers. Consequently, health and well-being are interrelated across generations, emphasizing the importance of promoting health and well-being across the life course. Therefore, this dissertation focuses on health and well-being across three generations – children, parents, and grandparents.[5]

Given the centrality of health and well-being to the productivity of single individuals, their families and society at large, it is crucial to better understand the factors that can influence health. These are many (e.g., genetics, environmental factors, health behavior, socio-economic status), and some are considered more proximal while others are more distal, but nonetheless important. While policies directly targeting health, such as reforms to the health insurance system, revisiting bans on drugs, and updated prevention measures, are usually assessed in relation to their health-related outcomes, policies affecting health indirectly are rarely examined from a health perspective. Indeed, a long line of research has shown, for example, that education can impact health. For example, several compulsory schooling reforms (intended to affect human capital accumulation) are shown to impact health (e.g., Fischer et al., 2013; Kemptner et al., 2011; Silles, 2009). Such indirect effects of social and family policies are the focus of this dissertation.

This dissertation adopts a broad perspective on health, and sees it as affected by a number of social and family policies which are not necessarily and explicitly framed as "health interventions." The aim of the thesis is to uncover many of the indirect effects that social and family policies can have on health and which often operate through intergenerational channels. Specifically, Chapters 2 and 5 deal with the effects of an increase in the early retirement age on women's health and associated healthcare costs. Chapter 3 analyzes the effect of grandparental childcare on parental and child outcomes, and Chapter 4 evaluates the impact of a daycare expansion on children's health(care consumption). Lastly, Chapter 6 contrasts the effects on child and maternal well-being and behavior with the program costs of a nursery program targeting parenting skills of disadvantaged families.

[5]Throughout this dissertation, the term "grandparents" is used for older individuals, i.e., individuals from the grandparents' generation. Chapter 3 specifically deals with grandparents, i.e., individuals that have grandchildren, while Chapters 2 and 5 look more generally at older individuals from age 59.

In Chapter 2, my co-authors and I study how an increase in the early retirement age for women from 60 to 63 impacts various health outcomes. To sustain the pension systems' financial stability in an aging society, the retirement age has been increasing across OECD countries. These policies aim to reduce the share of pension benefit recipients and increase the share of contributors, thereby almost universally affecting older adults. A plethora of studies show that raising the legal retirement age leads to later retirement (e.g., Atalay and Barrett, 2015; Geyer and Welteke, 2021; Staubli and Zweimüller, 2013), but the impact on health is yet unclear. Thus, understanding how prolonging working life affects individual health is critical to improving the health of older adults.

Apart from affecting health, changes in the retirement age may have unintended consequences, such as changes in care provision. Prolonging the working life imposes a time constraint on the affected people resulting in less informal care provision by the affected older individuals to other older adults (e.g., to the partner or parents, Fischer and Müller, 2020) and grandchildren (e.g., Backhaus and Barslund, 2021; Frimmel et al., 2020; Rupert and Zanella, 2018). In turn, a reduction in grandparental childcare provision due to a prolonged working life affects maternal labor market participation (e.g., Bratti et al., 2018). While the effects of grandparental childcare on grandparental and maternal labor supply are well-established, evidence on the effects on grandparental health and well-being is mixed (e.g., Arpino and Bordone, 2014; Danielsbacka et al., 2019). Moreover, the effects of grandparental care may go beyond grandparents themselves and also impact the care-receiving generation, i.e., parents and children. In Chapter 3, we investigate the intergenerational effects of grandparental care on parental well-being and child health, socio-emotional behavior, and school outcomes. Although a few studies in the economic literature analyze the effects of grandparental care on child outcomes (Ao et al., 2021; Del Boca et al., 2018), the evidence on parental well-being is scarce.[6]

A childcare option studied more extensively with respect to child and parental outcomes is publicly funded or highly subsidized daycare; however, evidence on the effects on child health is limited and ambiguous. In Chapter 4, I analyze the effects of early daycare attendance of children on their health(care consumption). Since the 19080s, most OECD countries have expanded publicly funded daycare provision - first for children aged three and older, and then, more recently, for children under three years. There are two main underlying reasons for expanding publicly funded daycare provision: First, offering daycare slots facilitates the reconciliation of childcare and employ-

[6] Chen and Zhang (2018) who analyze the causal effect of grandparental retirement on maternal well-being represent an exception.

ment. In essence, as mothers are still the primary caregiver, daycare availability allows mothers to expand their labor supply. A second policy goal of daycare expansions is to provide an educational environment for children, fostering child development and reducing social inequalities.

Numerous studies for Germany (e.g., Bauernschuster and Schlotter, 2015; Müller and Wrohlich, 2020)[7] as well as other countries (e.g., Baker et al., 2008; Berlinski and Galiani, 2007; Cascio, 2009; Gelbach, 2002; Havnes and Mogstad, 2011a; Nollenberger and Rodríguez-Planas, 2015) evaluate the effects of daycare on maternal employment.[8] Simultaneously, there is a wide field of empirical literature evaluating the effects of daycare on child development and – less often – on non-labor market outcomes of parents. The evidence on targeted programs – common in the Anglo-Saxon countries – that are usually aimed at children from socially disadvantaged backgrounds mostly highlights positive effects on child development including health.[9] In contrast, evidence on the effectiveness of universal daycare systems adopted by most continental European countries on child development is mixed. What is known is that children from disadvantaged families benefit disproportionately from daycare (see Cornelissen et al., 2018; Felfe and Lalive, 2018 for evidence from Germany and, e.g., Datta Gupta and Simonsen, 2010; Drange and Havnes, 2019; Felfe et al., 2015; Havnes and Mogstad, 2011b, 2015, for evidence from other countries).[10][11]

Evaluating the effect of daycare is particularly policy relevant as the majority of children attend a daycare center before they enter school due to the universal offer of daycare slots.[12] Additionally, although health in childhood is identified as one of the most important drivers of future educational achievements, alongside health outcomes and labor market success during adulthood (see, e.g., Carneiro et al., 2007; Currie, 2020; Currie and Stabile, 2006; Heckman et al., 2013; Heckman, 2007; Peet et al., 2015), the focus of studies evaluating the effects of daycare on child development lies on the formation of cognitive and socio-emotional skills. Thus, policies expanding

[7] See Spiess (2022) for a recent literature overview with a focus on Germany.

[8] Another widely studied aspect is the effect of day care provision on fertility (see Bauernschuster et al., 2016; Hank et al., 2004; Rindfuss et al., 2010, for studies on Germany and Norway, respectively).

[9] For example, Conti et al. (2016) and Heckman et al. (2010) provide evidence on the effects of two prominent US programs (Perry Preschool Program and Abecedarian Project) on child development.

[10] Although children from migrant and lower educated parents benefit disproportionately from attending daycare, there is an enrollment gap in the German daycare systems, i.e., children from non-migrant and high-educated parents are over-represented (Jessen et al., 2020).

[11] The effects of daycare on parental well-being is less investigated and shows mixed results (e.g., Baker and Milligan, 2008; Brodeur and Connolly, 2013; Herbst and Tekin, 2014; Kröll and Borck, 2013; Schmitz, 2020; Schober and Schmitt, 2017; Schober and Stahl, 2016).

[12] In 2022, in Germany 36% of children under three are enrolled in daycare while 91% of children aged three to six visit daycare (Destatis, 2022c).

daycare slots affect almost all children, calling for thorough evaluations of the effects on child health.

In the evaluation of social and family policies, next to the effectiveness, the resulting costs are of interest to policymakers. In the face of scarce public resources, efficiency studies are an important tool for policymakers and can help make investments in particular programs more compelling. In Chapters 5 and 6, I take a cost-perspective and analyze the impact of the retirement reform analyzed in Chapter 2 on healthcare costs (Chapter 5) and contrast the effectiveness and costs of a parenting program (Chapter 6).

Changes in the health status of individuals induced by raising the retirement age may also lead to changes in incurring healthcare costs. When an individual's health status improves, she likely demands less healthcare, thereby producing lower costs. The opposite happens when an individual's health status deteriorates. Thus, changes in the retirement age can lead, through changes in the health status, to indirect costs (e.g., changes in productivity) and direct costs through changes in the incurring healthcare costs. Therefore, in Chapter 5, we comprehensively evaluate the fiscal effects of the retirement reform by comparing the revenues (increased pension contributions) to the incurring healthcare costs.

In Chapter 6 of this dissertation, we analyze both program costs of a parenting program and its effectiveness with regard to maternal and child outcomes. Specifically, we evaluate the impact of a home-visitation program that aims to improve parenting skills and enhance knowledge on parental and child outcomes, contrasting these findings to the costs per child to run the program. In Chapters 3 and 4, I discuss the effects of two childcare settings – grandparental care and daycare – on parental and child health and well-being. However, it is not just the mode but also the quality of care that may affect child and parental outcomes. The literature emphasizing that the quality of daycare is at least equally important for child development as daycare attendance *per se*, is growing (see, e.g., Blanden et al., 2022; Kuger et al., 2019; Spiess, 2022, and references therein). Similarly, programs targeting parenting skills and knowledge – improving the quality of parental care – can lead to improvements in parental and child outcomes (see, Heckman and Mosso, 2014, for a literature overview). Addressing parenting skills and knowledge is particularly relevant for two reasons. First, parents still act as the main care actor for young children; for example, in Germany, almost 60% of children below three are only cared for by their parents (see Figure 3.2). Secondly, formal care is subject to strict quality regulations, thus, guaranteeing a minimum quality standard. However, the quality of informal care options (grandparents, parents) is more difficult to assess and may exhibit large differences between socio-economic groups. Thus, enhancing parenting skills and knowledge – particularly of parents from lower socio-economic backgrounds – can lead to improved parental and child outcomes for a large share of the population.

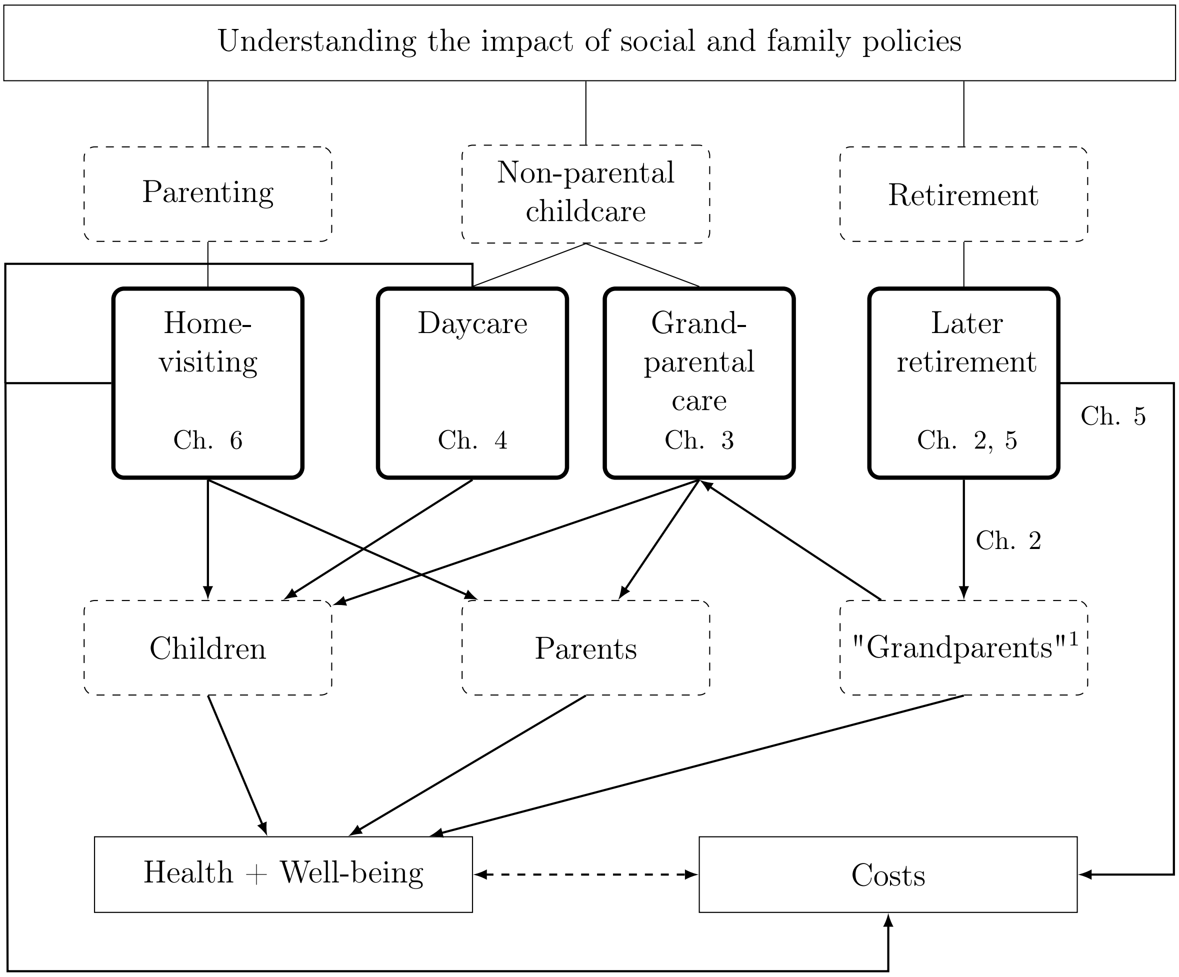
1.2 Overview and summary

This dissertation comprises five empirical papers in health and family economics. While each chapter is self-contained and constitutes independent contributions to the economic literature, the research questions are still closely linked, as outlined in the previous section. The chapters can broadly be categorized into two content-defined groups: **Chapters 2, 3, and 4** concern the effects of social and family policies on (grand-)parental and child health and well-being. **Chapters 5 and 6** take a cost perspective and contrast the effectiveness and costs of two social policies. Despite covering different topics, an overarching theme of the chapters of this dissertation is that they examine the (unintended) consequences of social and family policies on the health and well-being of three generations.

Figure 1.1 illustrates the connections between the Chapters. The rectangular boxes with the dashed frames represent the overarching political fields of action (parenting, childcare, retirement) and the generations under study (children, parents, grandparents). The bold-framed squared boxes depict the specific policy under study (home-visiting program, daycare, grandparental care, later retirement). Later retirement affects older individuals ("grandparents"), who, in turn, partly provide grandparental care. Grandparental care and daycare – the two non-parental childcare settings under study – affect both children and parents. Similarly, the home-visiting program designed to improve parenting quality impacts both children and parents. Overall, the chapters relate to two main outcomes of interest: health (including well-being) and costs. In all chapters (except for Chapter 5), the impact of the studied policy or childcare setting on health and well-being is directly measured. In some Chapters, I additionally take a cost perspective. In the following, I provide brief summaries of each chapter of this dissertation. The key points are also summarized in Table 1.1.

Chapter 2 explores the health effects of prolonged working life. Specifically, we analyze the causal effect of an increase in the early retirement age on official health diagnoses. Using a difference-in-differences approach, we exploit a sizable cohort-specific pension reform for women that raised the early retirement age from 60 to 63 years. The analysis is based on official records obtained via the National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung,

Figure 1.1: Connection between chapters



[1]The term "grandparents" represents older individuals, i.e., individuals from the grandparents generation. Chapter 3 specifically deals with grandparents, i.e., individuals who have grandchildren, while Chapters 2 and 5 focus on older individuals in general who are at least 59 years old.
Source: Own illustration.

KBV) covering all individuals insured by the public health system in Germany and including all certified diagnoses by practitioners. This enables us to gain a detailed understanding of the multi-dimensionality of these health effects. We study three dimensions of health: mental health (including mental and behavioral diagnoses), physical health (including metabolic/nutritional, circulatory/heart, and musculoskeletal diagnoses), and healthcare consumption. In the analysis, we differentiate between three age groups to cover potential anticipation effects (at age 59), direct effects (60–62 years), and post-employment effects (63–65 years). The empirical findings reflect the multi-dimensionality but allow for deriving two broader conclusions. First, we provide evidence that the increase in the retirement age negatively affects health outcomes as

Table 1.1: Overview and summary of the following chapters

	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
Title	The Effects of an Increase in the Retirement Age on Health – Evidence from Administrative Data	Does Grandparenting Pay off for the Next Generations? Intergenerational Effects of Grandparental Care	Age-specific Effects of Early Daycare on Children's Health	The Effects of an Increase in the Retirement Age on Health Care Costs – Evidence from Administrative Data	Costs and short-term effects of a home-visiting program in *BRISE* – first steps for a cost-effectiveness analysis
Research question	How does an increase in the early retirement age for women affect health?	What is the effect of grandparental care on child and parental outcomes?	How does early daycare attendance affect children's health?	How does an increase in the early retirement age for women affect healthcare costs?	What are the short-term costs and effects of a home-visiting program on parental and child outcomes?
Main finding	The prevalence of mental disorders, obesity and musculoskeletal diseases increases	Little or no effect on children and improvements in parental well-being	Substitution of illness spells of infectious diseases from elementary school age to the first years of daycare	Healthcare costs increase but the increase is small compared to the overall fiscal effects of the reform	Statistically insignificant effects across outcomes and generally lower costs compared to other programs
Data	KBV	Pairfam & SOEP	KBV, Destatis, RKI survstat, INKAR & SOEP	KBV	BRISE
Empirical approach	difference-in-differences& regression discontinuity design	Instrumental variable approach	difference-in-differences& event study	difference-in-differences	Instrumental variable approach combined with entropy balancing

Notes: KBV = Kassenärztliche Bundesvereinigung, Pairfam = Panel Analysis of Intimate Relationships and Family Dynamics, SOEP= Socio-economic Panel, RKI= Robert-Koch Institut, INKAR= Indikatoren und Karten zur Raum- und Stadtentwicklung, BRISE= Bremen Initiative zur Stärkung frühkindlicher Entwicklung.
Source: Own illustration.

the prevalence of several diagnoses, e.g., mental health, musculoskeletal diseases, and obesity, increases. Effect sizes vary between four and ten percent depending on the outcome and age group. Secondly, we do not find support for an improvement in health related to a prolonged working life. These findings are robust to sensitivity checks and do not change when correcting for multiple hypothesis testing. Our results imply that it is crucial for policymakers to consider health effects when considering future pension reforms.

In **Chapter 3**, we adopt a double-generation perspective and investigate the causal impact of grandparental childcare on children's health, socio-emotional behavior, school outcomes, and parental well-being. Based on two representative German panel data sets (Pairfam and SOEP) and exploiting arguably exogenous variation in geographical distance to grandparents, we analyze age-specific effects, taking into account alternative care modes. Specifically, we evaluate the impact of grandparental care on all children between zero and ten years and their parents, as well as separately for children who are in (full-time) daycare or school and children who are not. Our results suggest mainly null and, in a few cases, negative effects on children's outcomes. In case children three years and older are in full-time daycare or school and, in addition, regularly cared for by grandparents, they have more health and socio-emotional problems, in particular conduct problems, compared to children who are only half-day in daycare or school. In contrast, our results point to positive effects on parental satisfaction with the childcare situation and leisure. The effects for mothers correspond to an increase of 11 percent in satisfaction with the childcare situation and 14 percent in satisfaction with leisure, compared to the mean, although the results differ by child age. While the increase in paternal satisfaction with the childcare situation is, at 21 percent, even higher, we do not find an effect on paternal satisfaction with leisure. In addition, we conduct heterogeneity analysis employing causal forests (Wager and Athey, 2018) to understand the impact of grandparental care on specific subgroups. We provide evidence that grandparental care provided by younger and unhealthy grandparents negatively affects child health. Furthermore, while the effects on maternal satisfaction are driven by mothers who hold a university degree, the opposite picture is visible for fathers.

Chapter 4 aims to improve our understanding of the effects of early daycare attendance on children's health. Despite the relevance of child health for child development and later life success, the effect of early daycare attendance on health has received little attention in the economic literature. In this study, I investigate the impact of a large daycare expansion in Germany on children's age-specific mental and physical health outcomes. Based on the same administrative health records used in **Chapter 2**, which cover 90% of the German population over ten years, I exploit exogenous variation in daycare attendance, taking advantage of a large-scale daycare expansion. Specifically, I employ difference-in-differences and event study approaches, exploiting temporal and spatial variation in the expansion speed of daycare slots. The results provide evidence for the substitution of illness spells from elementary school to the first years of daycare. Concretely, I find that early daycare attendance increases the prevalence of respiratory and infectious diseases and healthcare consumption when entering daycare (1–2 years) by 5–6 percent. At elementary school age (6–10 years), the prevalence decreases by similar magnitudes. I do not find evidence for an effect of daycare attendance on mental disorders, obesity, injuries, vision problems, or healthcare costs. I then conduct subsample analyses to understand potential heterogeneity in the effects. While there is no difference in the effects between boys and girls, the results indicate more pronounced effects for children from deprived areas for infectious diseases, earlier detection of vision problems and a decrease in the prevalence of obesity in these children. Lastly, I discuss the implications of the results reflecting on potential long-term health effects, spill-over effects on siblings and parents, duration of illness spells for different age groups, and sickness absence of children in daycare or school, concluding that there is no evidence that changing the timing of infections to earlier years leads to detrimental effects that would challenge the daycare entry age of children from a health perspective.

Chapter 5 builds on the analysis of **Chapter 2** by estimating the effect of the increase in the early retirement age for women on healthcare costs. Thus, in this Chapter, we take a cost perspective and analyze how the health effects translate into costs relying on the same data as in **Chapter 2**. Our results show that the reform led to a significant increase in outpatient care costs by about 2.9% (about 16 euros per individual) in the age group directly affected by the increase in the retirement age (60-62). Moreover, we also find expectation effects for women aged 59 and indirect post-employment effects for women between 63 and 65. We further show that the cost increase is mainly driven by the utilization of the following specialist groups: Ophthalmologists, general practitioners (GPs), oral and maxillofacial surgery, neurology, orthopedics, and radiology. The absolute effect is largest for GPs (about 3.5 euros), thereby contributing about 25% to the increase in the overall costs. While the effects are significant and meaningful on the individual level, we show that the increase in healthcare costs is modest relative to the positive fiscal effects of the pension reform. Specifically, we estimate an aggregate increase in health costs of about 7.7 million euros for women aged 60-62 born in 1952. The corresponding estimate of the net effects of the pension reform for the tax and transfer system, including social security, amounts to about 4 billion euros (Geyer et al., 2020). Thus, the increased healthcare costs amount to only about 2% of the generated revenues.

Chapter 6 of this dissertation deals with the costs and effects of a parenting program targeting the parenting skills of parents from disadvantaged backgrounds. Specifically, we evaluate the short-term effectiveness and costs of the home-visiting program *Pro Kind*, which targets parenting styles and parental behavior during pregnancy and early childhood. It entails bi-weekly home visits starting prenatally and lasting until children turn two. The main interest of this Chapter is to understand whether *Pro Kind* already had a significant impact on mother and child outcomes during the first seven months of the children's lives. *Pro Kind* is the first program within a systematic chain of home- and center-based preschool interventions established to support disadvantaged families from pregnancy through school entry under the Bremen Initiative to Foster Early Childhood Development ("Bremen Initiative zur Stärkung frühkindlicher Entwicklung"

– *BRISE*). In the first step, we analyze the program's effectiveness on a range of child and maternal outcomes (e.g., child development, maternal smoking, maternal alcohol consumption, and maternal mental health). We make use of an access and information treatment randomly assigned to *BRISE*-families to predict the families' *Pro Kind* participation probability. Our results, based on an instrumental variable approach combined with entropy balancing (Hainmueller, 2012), show no significant short-term effects of the program on child and maternal outcomes. Secondly, we analyze the program costs drawing on comprehensive, self-collected data from a yearly cost survey following the ingredient method (Levin and McEwan, 2000). The resulting micro cost data set covers all costs related to the program implementation. We then proceed to compare the overall cost per child per year to other well-established programs and find that, *Pro Kind* is less costly than most comparable programs.

Finally, **Chapter 7** concludes, summarizes the main policy implications of all chapters, and critically discusses the limitations and scope for future research.

1.3 Contribution

This dissertation makes several important contributions to the literature. While each Chapter makes an independent contribution to the economic literature – discussed in detail in the respective Chapters – I will outline six main contributions that this dissertation makes in general and that go beyond the individual Chapters: (1) Emphasizing the *sensitivity of health to a variety of factors and policies*, (2) assessing the *multi-dimensionality of health*, (3) evaluating *age-specific effects*, (4) adding a *cost perspective*, (5) drawing on a *variety of data sources*, and finally, (6) all Chapters contribute to the understanding of the *German context*.[13]

First, this dissertation presents evidence for the *sensitivity of health to a variety of factors and policies* related to the family and labor market. Specifically, I show that health is indirectly affected by several social and family policies that do not have explicit health objectives. The social and family policies under study cover various aspects, ranging from enhancing the parenting skills of parents of newborns, to childcare modes for children between one and ten years to a retirement reform. Thereby, the results of this dissertation emphasize that all kinds of social and family policies may have unintended consequences for health and well-being.

[13] All contributions are content- or data-related. I will not discuss *methodological*-related contributions, as I apply in all Chapters well-established quasi-experimental approaches (e.g., difference-in-differences, regression discontinuity approach, instrumental variable approach) to estimate causal effects. Thus, I do not consider the application of these methods as contributions but rather as tools to answer my research questions.

Second, this dissertation highlights the *multi-dimensionality of health.* While generally, health is responsive to distal policies, not all health dimensions respond in the same way. To generate a comprehensive understanding of the impact of the social or family policy on health and well-being, all Chapters consider an extensive range of outcomes. For example, in **Chapter 2**, I find adverse effects of increasing the retirement age on health for some diagnoses and null effects for others. Similarly, in **Chapter 4**, I provide evidence that early daycare attendance affects the prevalence of communicable diseases but does not affect other diseases, such as injuries and mental disorders. Additionally, in two chapters (**Chapters 3 and 6**), I consider the health and well-being outcomes of two generations. My results support this double-generation approach, as effects vary across generations. For example, in **Chapter 3**, I generally find null or negative effects of grandparental care on child health and socio-emotional skills, while there are positive effects on parental well-being.

Third, as health effects are produced along the entire life course, this dissertation deals with different generations and assesses *age-specific effects.* The results show that different age stages are affected by the social and family policies, namely children, their parents, and older adults, emphasizing the embeddedness of health within social and family policies at all life stages. Moreover, I am able to show that even within life stages, health effects are often age-specific. Specifically, in each Chapter, I differentiate between age groups within one generation. Age-specific effects are particularly interesting when analyzing the effects of certain life events and transitions, such as entering retirement, as health and well-being may react differently before, during, and after the event. For example, in **Chapters 2 and 5**, we study the age-specific effects of an increase in the early retirement age on women's health(costs). Specifically, we compare the health outcomes of women at age 59, when neither of the cohorts is eligible for retirement (expectations effects), at age 60–62 years when only one cohort is eligible for retirement (direct effects), and at age 63–65 years when both cohorts are eligible for retirement (post-employment effects). Similarly, in **Chapters 3 and 4**, we evaluate the effects of two childcare settings (grandparental care and daycare) for children between zero and ten years. In both chapters, I analyze the effects for the pooled age group (age 1–10 years) as well as for delicate age groups. In **Chapter 6**, we differentiate between age groups on an even finer level by studying the effects of a parenting program on three and seven months old children and their mothers. The results in all Chapters depict significant heterogeneity across age groups, emphasizing the importance of studying not only different generations but also different age groups within the generations.

Fourth, as health can be good or poor, it can create both benefits and costs. This dissertation maintains the focus on both and is thus going beyond usual effectiveness studies by taking a *cost perspective*. Specifically, **Chapters 4 and 5** analyze the impact of the social policy under study on healthcare costs. Thereby, we translate the impact on the individual health status into direct costs for the healthcare system. While the health effects most certainly also imply indirect costs (e.g., through changes in productivity), assessing the direct costs is a critical contribution to policy evaluations. **Chapter 6** takes a different cost perspective by contrasting the effects on children and mothers and the costs of a parenting program. The literature on cost-efficiency studies of parenting programs is scarce (especially in the European context), but an essential tool for policymakers to decide on the introduction or continuation of programs.

Making these contributions – adding to the knowledge on the health effects of social and family policies – required drawing on a *variety of data sources*. Hence, this dissertation indirectly makes the contribution of uncovering a wealth of data sources, choosing and combining the right data sets to answer relevant research questions. Specifically, three Chapters use administrative health records collected by the German National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung, KBV). These data include all publicly health-insured individuals in Germany, which amounts to about 90% of the population. Using this data has several advantages: First, as it covers almost the whole population, the results have a high external (results are not specific to a selective subpopulation) and internal validity (large sample size allows the application of methods that generate unbiased estimates). Second, it includes objective and detailed health outcomes. Survey data usually include subjective health measures, i.e., assessed by non-healthcare professionals, which can be subject to biases (e.g., Bound et al., 2001). Additionally, health measures obtained in survey data are often broad and unspecific (e.g., the general health status). Detailed diagnosis data allows for studying a range of outcomes, which has benefits as outlined above. In **Chapter 4**, I combine the KBV data with data from the Federal Statistical Office on regional daycare coverage rates, data on the regional swine flu incidence obtained from the RKI, and regional INKAR data on county characteristics (e.g., average income). I supplement the analysis with an analysis using the Socio-economic Panel (SOEP), as parts of the question (spill-over effect of daycare attendance on parents) cannot be answered with the administrative data due to a missing link between children and parents. This data limitation of the KBV data (and many other administrative data sources, at least in Germany) highlights survey data's benefits in answering certain questions. In many panel survey data, such as the SOEP and Pairfam, it is possible to link generations, allowing to answer research questions that require an intergenerational perspective. Another advantage of survey data is that they contain information that is not included in administrative data, such as subjective well-being. Therefore, in **Chapters 3 and 6**, where we consider both child and parental outcomes, we rely on survey data, namely SOEP, Pairfam, and the data collected within the BRISE project.

Lastly, the empirical analyses in all Chapters are based on the *German context* and therefore contribute to the better understanding of the social welfare system in Germany and countries with similar institutional settings as a whole. Germany is a particularly interesting case study for several reasons. First, Germany is characterized by a generous social welfare system, including a public pension system covering about 90% of the workforce, a universal public health insurance system covering about 90% of the population[14], and a universal daycare system that is highly subsidized. While the universal pension and health insurance systems have been in place for a long time, publicly funded daycare provision is a more recent trend. Thus, maternal labor market participation is traditionally lower than in many other OECD countries (e.g., the US, the UK, or the Nordic countries), with an increasing trend over the past decades. Thus, studying the effects of childcare settings in Germany is particularly interesting, as most previous studies are based in the US, UK, or the Nordic countries, where the institutional settings and social norms are different. The universal health insurance system also provides an interesting setting for studying the impact of life transitions (e.g., into retirement or daycare) on health, as it allows isolating the "pure health effects" from effects caused by the availability of healthcare. For example, in the US, healthcare insurance is often attached to employment, such that entering retirement leads to changes in the insurance status of individuals. Overall, providing evidence on the effects of social policies in the German context advances the understanding of the potential scope of such policies in Germany and other countries facing a similar institutional setting.

[14]Health insurance is mandatory. Individuals who are not insured via a public health insurance fund have to be insured by a private insurer.

CHAPTER 2

The Effects of an Increase in the Retirement Age on Health – Evidence from Administrative Data[1]

2.1 Introduction

Aging populations present immense challenges for public pension systems due to growing numbers of beneficiaries and declining numbers of contributors. To sustain the systems' financial stability, policy makers across the OECD have introduced pension reforms which raised retirement ages. While postponing retirement has the potential to increase pension contributions and to reduce the share of pension benefit recipients, a prolonged working life might also have consequences for the health of individuals. Thus, to understand and to assess the overall impact of changes to the pension system, it is crucial to quantify and fully understand the health implications of pension reforms.

In this paper, we study the health effects of an increase in the retirement age using official data on certified diagnoses by practitioners based on the International Classification of Diseases (ICD-10) for the period from 2009 to 2018. The detailed information on specific diagnoses and groups of diseases allows us to analyze the implications for health outcomes in a multi-dimensional way. This detailed analysis is important since

[1]This chapter is joint work with Johannes Geyer (DIW Berlin), Peter Haan (DIW Berlin and Freie Universität Berlin) and Anna Hammerschmid (former DIW Berlin). We are grateful to the National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung, KBV) for data access and for their excellent support. We further thank Lena Janys, Adam Lederer, and Marius Opstrup Morthorst, as well as the participants at the Essen Health conference 2020, The Econometric Society/Bocconi University World Congress 2020, and at internal seminars at DIW Berlin. Moreover, Peter Haan gratefully acknowledges funding from the German Science Foundation through the CRC/TRR190 (Project number 280092119) and Project HA5526/4-2 and funding from JPI More years better lives through PENSINEQ. We also thank four anonymous referees and the editor of this issue of The Journal of the Economics of Ageing, Alfonso Sousa-Poza, for valuable comments and suggestions. This chapter contains the version submitted to the Journal of the Economics of Ageing after the second round of revision.

broader health measures which have been used in most of the previous studies, might disguise potentially negative or positive implications for different health dimensions.

To identify the causal effect of an increase in the retirement age on diagnoses, we exploit a sizable and cohort-specific pension reform which was implemented in 1999. The reform abolished an early retirement program for women born in 1952 and after[2] and thereby effectively increased the early retirement age (ERA) for women from age 60 to at least 63. It provides a clean quasi-experimental setting as it induces a substantial discontinuity in retirement ages for two adjacent cohorts (women born in 1951 versus women born in 1952). Using the same variation, Geyer and Welteke (2021) and Geyer et al. (2020), analyze the employment effects as well as distributional consequences of the pension reform and show that the reform led to substantial individual labor market responses, including increased employment between age 60 and 62. Moreover, Etgeton et al. (2021) show that the reform had negative effects on private savings.[3] Using data covering 2009 through 2018, we can consistently analyze the health effects for women aged 59, i.e. before the reform had a direct effect on employment (age-59-effects), for women aged 60–62 (main effects) and for women aged 63–65, which we define as post employment period.

In the main analysis we use a Difference-in-Differences (DiD) design. The medical and demographic literature documents that health outcomes are correlated with month of birth as well as with cohort effects (e.g., Boland et al., 2015; Doblhammer and Vaupel, 2001). Therefore, it is crucial to account for cohort and seasonality (month of birth) effects to isolate the causal effect of the pension reform on health. Similar to Schönberg and Ludsteck (2014), we define a treatment group (women born between October 1951 and March 1952) and a control group (women born between October 1950 and March 1951) which captures cohort and seasonality effects.

In the analysis, we focus on three dimensions of health: mental health, physical health, and healthcare consumption. Within these dimensions, we concentrate on groups of diseases that are most likely affected by lifestyle choices and that have been used in existing studies on the link between health and retirement. Within these groups, we select the diagnoses most frequently causing rehabilitation treatments prescribed

[2]The majority of previous studies on the link between health and retirement use age discontinuities in the retirement age to instrument the individual's retirement status (see van Ours and Picchio (2020) for an overview of methodologies of previous studies). Only a few studies exploit direct variation from pension reforms (e.g., Bloemen et al., 2017; Charles, 2004; Etgeton and Hammerschmid, 2019; Grip et al., 2012; Kuhn et al., 2019).

[3]To date, few other studies exploit variation from the 1999 pension reform: e.g., Gohl et al. (2020) use the reform to test the human capital theory and Fischer and Müller (2020) analyzes its impact on informal care provision. Moreover, Etgeton and Hammerschmid (2019) study the effects of retirement on self-reported health, in particular across educational groups, using a two-sample-2SLS approach.

by the pension insurance in the application process of invalidity benefits ("Erwerbsminderungsrente"). More precisely, we analyze the impact of the increase in the retirement age on mood (affective) disorders and on neurotic, stress-related, and somatoform disorders (hereafter: stress-related diseases) to assess the effects on mental health. For the physical health dimension, we consider the group of metabolic and cardiovascular diseases (diabetes mellitus, obesity, ischaemic heart diseases, and cerebrovascular diseases (strokes)) as well as the group of musculoskeletal diseases (arthrosis and other dorsopathies). In addition, we study hypertension since this is the most common physical disease within our sample, but is not captured using the rehabilitation criterion. To estimate the impact on healthcare consumption, we examine the annual number of treatment cases.

Our empirical findings provide evidence that the increase in the retirement age has a negative effect on health outcomes as the prevalence of several diagnoses, e.g. mental health, musculoskeletal diseases, and obesity, increases. In contrast, we do not find support for an improvement in health related to a prolonged working life since there is no significant evidence for a reduction in the prevalence of any health outcome we consider. These findings are robust to sensitivity checks, and do not change when correcting for multiple hypothesis testing. Further, placebo tests provide empirical support for the identification assumptions of the DiD.

In particular, we find that the pension reform increased the prevalence of both mental diseases in 60–62 year old women. The effect amounts to 3.6 percent for stress-related diseases and to 4.8 percent for mood disorders relative to the respective pre-treatment means. The effects for 59 year old women are of similar significance and about twice as large. Within the physical health dimension, our results suggest that raising the retirement age increases the prevalence of dorsopathies, arthrosis and obesity at ages 60–62 years as well as 59 years. For other physical health outcomes, our results are less clear but, as mentioned above, we do not find significant evidence of an improvement in physical health in response to the reform. Furthermore, we find significant effects of the reform on healthcare consumption for 59 year olds. Overall, our findings reflect the multi-dimensionality of health but allow us to conclude that the reform had negative and significant effects on some health outcomes and did not have positive and significant effects on any of the considered health outcomes. Additional analyses on post-employment effects suggest that the majority of the effects persist into retirement (at age 63–65), but effect sizes are smaller compared to the direct effects on 60–62 year old women.

Literature

The existing literature on the health effects of retirement and pension reforms can be divided into four strands: Studies using survey data and exploring effects of retirement on i) mental health or ii) physical or general health, and studies using administrative data considering iii) mortality or iv) healthcare usage or diagnoses as outcome variables. We discuss the relation of our paper to these four strands in the following:[4]

Survey data: Mental health
A number of studies find positive effects of retirement on mental health (e.g., Atalay and Barrett, 2014; Belloni et al., 2016; Charles, 2004; Eibich, 2015; Gorry et al., 2018; Grip et al., 2012; Leimer and Van Ewijk, 2022; van Ours and Picchio, 2020). Atalay and Barrett (2014), for example, exploit variation of a pension reform for women in Australia and find positive effects of retirement on mental health. They emphasize that the effects can mostly be attributed to a reduction in mood disorders. Eibich (2015) uses data from the German Socio-economic Panel (SOEP) and a Regression Discontinuity Design (RDD) exploiting age thresholds in the German pension system. He also finds positive effects of retirement on mental health and explains this by a reduction in work-related stress and more frequent exercise (cf., Celidoni and Rebba, 2017). Applying a similar methodology van Ours and Picchio (2020) find heterogeneous effects for the Netherlands. They find positive effects of retirement on the mental health of men and their partners but no effects for women or singles.

In contrast, there are also studies showing no, if not negative, effects of retirement on mental health (e.g., Atalay et al., 2019; Heller-Sahlgren, 2017; Mazzonna and Peracchi, 2017; Rohwedder and Willis, 2010). For example, Heller-Sahlgren (2017) conducts a cross-country analysis using the Survey of Health, Ageing and Retirement in Europe (SHARE) and employs an RDD approach. He finds no effects on mental health in the short-run but a large and negative long-run impact. Similarly, Rohwedder and Willis (2010) find negative effects on cognitive abilities in a cross-national study in the US and Europe. These results are also supported by Mazzonna and Peracchi (2017), who find a decline in cognitive abilities following retirement for most workers using SHARE data. Atalay et al. (2019) find a negative but modest effect on cognition, the effect is larger for men than for women.

Survey data: Physical and general health
The relationship between physical or general health and retirement is also ambiguous in the literature. Coe and Zamarro (2011) and Gorry et al. (2018) find positive ef-

[4]For a more detailed overview of the literature please refer to e.g., Garrouste and Perdrix (2022) or van der Heide et al. (2013).

fects of retirement on self-reported health status in Europe using SHARE data. Shai (2018) reports similar findings for Israel. Leimer and Van Ewijk (2022) uses SHARE data and reports a reduction in mobility limitations and the number of limitations in activities of daily living along with an increase in maximum grip strength following retirement. Close to our study, in particular in terms of the same reform being used for identification, is Etgeton and Hammerschmid (2019). They focus on the effects of retirement on broad, self-reported health, in particular across educational groups, based on SOEP and SHARE data. Using a two-sample 2SLS approach, they identify the impact of retirement on health using the 1999 pension reform in Germany. Their findings point toward non-detrimental general health effects of retirement, with less educated women benefiting more than the average.[5] In addition to positive effects on mental health, Atalay and Barrett (2014) also find positive effects on physical health, namely on hypertension, migraine, back pain, and disc disorders for women in Australia. These positive effects are in line with studies that show that retirement leads to changes in lifestyle habits such as increases in physical activity and sleep time and a reduction in drinking (e.g., Kämpfen and Maurer, 2016; Motegi et al., 2016).

Negative effects of retirement on physical health are found, for example, by Godard (2016) (increase in BMI with SHARE data), Behncke (2012) and Pedron et al. (2020). Specifically, Behncke (2012) discover an increase in risk of being diagnosed with a chronic condition and an increase in risk of developing a cardiovascular disease in the UK following retirement. Similarly, Pedron et al. (2020) analyze the KORA cohort study including older individuals in southern Germany making use of an RDD design exploiting age thresholds for pension eligibility. They document increases in the BMI among early retirees and increases in total cholesterol/HDL quotient in regular retirees.

Examples of studies assessing the effect of retirement on healthcare consumption are Zhang et al. (2018) for China, Eibich (2015) for Germany and Eibich and Goldzahl (2021), Coe and Zamarro (2015) and Lucifora and Vigani (2018) for various European countries. While Zhang et al. and Lucifora and Vigani report increased healthcare utilization following retirement, others provide evidence for a decrease in both hospitalization (Eibich, 2015) and number of doctor visits (Coe and Zamarro, 2015; Eibich, 2015) as well as reduced preventive care usage, particularly a reduction in mammography (Eibich and Goldzahl, 2021).

The reasons for the discrepancies in the literature are not comprehensively and systematically studied yet, but contributing factors seem to be, for instance, differences

[5] Also the results of Grøtting and Lillebø (2020) provide evidence for a positive effect of retirement on physical health especially for individuals with low socioeconomic status based on Norwegian survey data.

in empirical methods, data sources, pension systems, healthcare systems, effect heterogeneity in sub-populations, and differing outcome variables (Nishimura et al., 2018; Pilipiec et al., 2020).[6] Furthermore, heterogeneity in the effects of retirement on different health dimensions could potentially also contribute to explaining the contradictory results. There is ex-ante no reason to believe that the effects of retirement (reforms) on different health dimensions are indeed homogeneous and go into the same direction. Some aspects of mental or physical health may be positively affected whereas others may be negatively affected.

Administrative data: Mortality

Analyses using detailed administrative data including objective health measures have the potential to explore this issue. So far, only a small number of studies use this kind of data. Examples of studies looking at the effect of retirement on mortality are Kuhn et al. (2019), who find negative effects for Austrian men, Fitzpatrick and Moore (2018) for the US, and Brockmann et al. (2009) in the German context. Brockmann et al. (2009) use German health insurance data from one specific health insurance fund and find heterogeneous effects across individuals with good and poor health. Healthy people benefit from retirement while individuals with poor health tend to have decreased life expectancy following early retirement. In contrast, Hallberg et al. (2015) use a pension reform for military officers that decreased the retirement age from 60 to 55 in Sweden. They find support that early retirement leads to a reduction in mortality. Hernaes et al. (2013) find no effect of a series of retirement reforms that reduced the retirement age on mortality in Norway.

It is important to note that death is a specific and extreme outcome. Mortality rates are rather low around retirement age. Potential effects on mortality might only establish later in the long run. Thus, it is difficult to estimate mortality effects of recent pension reforms, such as the 1999 reform studied in this paper.

Administrative data: Health care consumption and diagnoses

Studies using administrative data and considering health outcomes other than mortality or healthcare consumption are less common; these mostly find positive effects of retirement on health. The following studies are closely related to our study:

Kuusi et al. (2020) use Finish registry data (a random sample covering 11% of the population) and an IV approach to assess the effect of retirement on mental health and physical health. They measure mental health with antidepressant purchases and physical health by hospital visits associated with cardiovascular or musculoskeletal diseases.

[6] Nishimura et al. (2018) show that the choice of empirical method plays a key role in explaining why estimated results differ across studies.

They find substantial positive effects on mental health and small effects on physical health. Similarly, Nielsen (2019) uses Danish full population data to assess the effect of retirement on general practitioner (GP) visits, hospitalization, comorbidities, and mortality using IV and RDD approaches. He finds a reduction in GP visits and hospitalization following the reform, but no effect on comorbidities and mortality. Hagen (2018) conducts a similar study in Sweden but does not find an impact of retirement on health. He uses Swedish data for women in the public sector to estimate the effect of a pension reform on drug prescriptions, hospitalizations, mortality, and cause-specific health indices in a DiD framework. There are only a few studies outside the Nordic countries relying on administrative data (e.g., Bíró, 2016; Bíró and Elek, 2018; Frimmel et al., 2020; Horner and Cullen, 2016; Perdrix, 2021; Rose, 2020). Horner and Cullen (2016) use administrative data from the US on a specific group, manufacturing workers in an aluminum production company, to evaluate the impact of retirement on hypertension, diabetes, asthma, arthritis, and major depression. They find a reduction in asthma following retirement but no effects on the other outcome variables. Frimmel et al. (2020) study the effect of two Austrian pension reforms on individual inpatient and outpatient healthcare utilization in Austria and find that retirement decreases service utilization and healthcare expenditure. In contrast, Bíró (2016) documents increased healthcare consumption for pensioners while other scholars provide evidence for decreases in outpatient care, inpatient care and prescribed pharmaceutical expenditures (Bíró and Elek, 2018) and doctor visits particularly GP visits (Perdrix, 2021). Rose (2020) uses a combination of administrative and survey data from the UK to study a variety of outcomes: She generally finds a positive association between retirement and health, e.g. an increase in self-reported health, a decrease in long-term ailments, lower pulses, more sleep and generally an improvement in healthy behaviors (e.g., reduced smoking and drinking). However, she does not find retirement to impact cognition, mental health, healthcare utilization and mortality.

Our paper extends the literature in several ways. First, we study a major pension reform that led to a substantial increase of the retirement age of three years. Second, our study is based on unique administrative health records that cover almost the whole German population. Moreover, the data include all recorded diagnoses in outpatient care during the observation period. Thus, in contrast to most of the previous studies we can study the multi-dimensionality of health effects for a very general population. Third, we provide evidence that effects from increasing the retirement age are not bound to the affected age group. Instead, increasing the retirement age implies expectation effects (effects for the age group before reaching the retirement age) and the effects persist into the post-employment period. Finally, in contrast to the previous literature which mainly focuses on men we provide evidence for the health effects of women.

The remainder of the paper is structured as follows: Section 2.2 describes the institutional background in Germany. In section 2.3, we give an overview over the data. The empirical strategy is explained in section 2.4 and, in section 2.5, we present the results and provide several robustness checks. Finally, section 2.6 concludes.

2.2 Institutional background - Pension system

To establish the institutional setting of the analysis, we provide an overview on the relevant institutions of the German pension system[7] and discuss the 1999 pension reform, which induced an exogenous increase in the early retirement age for women born in 1952 and after.

The public pension system in Germany covers about 90% of the workforce.[8] Pension benefits account for about two-thirds of gross income of the elderly. It includes old-age pensions, disability pensions, and survivors' benefits. The system is financed by a pay-as-you-go (PAYG) scheme and has a strong contributory link. The calculation of pension benefits is based on a points system and depends on the entire working history.[9] The statutory pension age (SRA) was 65 for cohorts born before 1947. It is stepwisely raised to age 67 and fully phased in for all cohorts born in 1964 or later. For the 1951 cohort, the SRA was 65 and 5 months, for those born in 1952 it was 65 and 6 months. People qualify for this regular old-age pension after five years of pension contributions.

Retirement before the SRA (with permanent deductions) is possible under certain conditions.[10] There are four alternative pathways to claiming early retirement benefits: the *pension for women*, the disability pension, the pension for the long-term insured, and the pension after unemployment or after partial retirement. There is a fifth option, invalidity benefits ("Erwerbsminderungsrente"), for people with severe health problems who are not able to work more than three hours a day.[11] In general, the calculation of pension benefits does not vary between these alternatives, whereas eligibility criteria

[7]For a more general description of the German pension system, see the German country profile by the OECD available at `http://oe.cd/pag`.

[8]There are a few exemptions from compulsory insurance: civil servants have a separate tax-financed, non-contributory scheme and most of the self-employed are not compulsory insured.

[9]People also acquire pension entitlements during short-term unemployment, for childcare, and for providing elderly care.

[10]There is no change to public health insurance coverage when starting to draw retirement benefits.

[11]People who are able to work more than three hours a day but less than six are eligible for partial invalidity benefits. These benefits are available before the age of 60.

differ.[12] The 1999 reform abolished the *pension for women* for cohorts born in 1952 and after. Effectively, the reform raised the ERA for most women from 60 to 63, which implies an extension of the working life of three years. The eligibility criteria of the *pension for women* were: (i) at least 15 years of pension insurance contributions; and (ii) at least 10 years of pension insurance contributions after the age of 40. According to Geyer and Welteke (2021), about 60% of all women born in 1951 were eligible for the old-age pension for women.[13]

Geyer and Welteke (2021) and Geyer et al. (2020) evaluate the labor market effects of the 1999 pension reform. Several findings of these studies are relevant for the subsequent empirical analysis. Most importantly, the increase in the ERA has sizable labor market effects: retirement rates of eligible women aged between 60 and 62 decreased by about 30 percentage points. At the same time, employment rates increased by about 15 percentage points (pre-reform mean 54%). Inactivity and unemployment increased by about 11 percentage points (pre-reform mean 12%). Moreover, the employment effect results almost entirely from women staying longer in the respective labor market status; there is no significant evidence that the unemployed make more transitions into employment. Unfortunately, we cannot identify labor market effects with our data, and thus we only can estimate the average effect and can not separately estimate the health effects for women in employment or in unemployment or inactivity. Further, Geyer and Welteke (2021) document that the pension reform had no significant effect on labor market activity before the age of 60 and they show that the pension reform did not lead to substitution effects into other health-related early retirement pathways (disability pension or invalidity benefits). The labor market effects of the pension reform are important for the interpretation of our results since the health effect we study can be linked directly to the changes in employment induced by the reform.

2.3 Data

For the analysis, we use administrative data covering the years 2009-2018, collected by all public health insurance funds in Germany.[14] In the data, physicians record a standardized diagnosis for each claim in order to be reimbursed by the health insurance.

[12]For more details see Geyer et al. (2020).

[13]In our data we cannot identify if women are eligible for the pension reform. Therefore, in the empirical analysis we estimate the intend to treat effect for all women. In the conclusion we add a back of the envelope calculation of the average treatment effect.

[14]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 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.[15] Only individuals with earnings exceeding a certain threshold[16] 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.[17]

With the data, in principle we can focus on women born between 1948–1953. However, a major school reform affects many women born after 1952, therefore in the empirical analysis we consider only cohorts 1948–1952.[18] This allows us to construct a control group (women born late in 1950 and early in 1951) in addition to the group of women around the cutoff date of the pension reform (women born late in 1951 and early in 1952). We have access to data covering 2009 through 2018, thus we can consistently analyze the health effects for women aged 59, i.e., before the reform had a direct effect on employment (age-59-effects), for women aged 60–62 (main effects) and for women aged 63–65, which we define as post employment period. As mentioned above women born in 1952 or later can enter retirement at age 63.

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, we construct a balanced sample with yearly information for all publicly insured individuals. First, we create variables indicating whether an outcome, for example diabetes, was diagnosed or not in a specific period. Secondly, we aggregate the data to a yearly level such that each patient appears only once per year. Finally, we 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 our outcome variables is analogous to van den Berg and Siflinger (2022). Thus, in the bal-

[15]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 (2019b).

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

[17]Importantly, similar rules apply for the eligibility of the public health and public pension insurance. Individuals with a private health insurance, e.g., civil servants and the self-employed, have additional private pension plans that were not affected by the pension reform.

[18]Regional schooling reforms in western Germany raised compulsory schooling from 8 to 9 years. Four large federal states changed compulsory schooling within cohort 1953. The reform had positive effects on health outcomes (Kemptner et al., 2011).

anced panel each patient appears between the first and last observed year every year.[19] The final data set includes about 500,000 women per birth cohort resulting in 2.5 million women overall. Women who did not receive any outpatient care during the 10 year observation period are not included in our sample. However, RKI (2010b) states, that 90% of women receive outpatient care at least once per year. Thus, given that we observe individuals over 10 years, the share of women not receiving any outpatient care should be negligible.[20] The data only includes few demographic characteristics, such as age and region.

Instead of estimating the effect for about 70,000 different diagnoses categorized by the ICD-10 codes, we use clear criteria to select the relevant health outcomes. Specifically, we concentrate on groups of diseases that are most likely affected by lifestyle choices and are used in the existing literature on the link between health and retirement. Within these groups, we select the diagnoses that most frequently caused rehabilitation measures prescribed by the German pension insurance for our age group.[21] In addition, we study hypertension since this is the most common disease within our sample and is not captured using the rehabilitation criterion.

Specifically, we define the following groups:

- Mental and behavioral disorders (ICD-10)

    F30-F39: Mood (affective) disorders

    F40-F48: Neurotic, stress-related and somatoform disorders (stress-related diseases)

- Endocrine, nutritional and metabolic diseases and diseases of the circulatory system (cardiovascular) (ICD-10)

    E10-E14: Diabetes mellitus

[19]Note, we do not impute information before the first year of observation or after the last year of observation.

[20]In the balanced sample, 79% of individuals appear every year. There are several explanation why the remaining 21% of the sample do not appear in every year. First, patients who die or leave the public health insurance system (e.g. move abroad or move to private insurance) leave our sample. Second, the construction of the anonymized patient ID is based on information such as name and birth date of the patient (that we do not observe). It happens, that patients have multiple IDs if at any doctor visit some information (e.g. name) is wrongly documented. Thus these IDs only appear once. The attrition is no threat to our identification strategy as it affects the treatment and control group in the same way. Moreover our identification strategy does not rely on multiple observations of the same individual in different years.

[21]Employees can receive medical rehabilitation benefits if their earning capacity is at considerable risk or already reduced. The goal is that individuals recover such that they can return to the labor market and do not need invalidity benefits. We selected the diseases that were responsible for at least 20% of the prescription cases within a group of diseases. The list is accessible at `https://statistik-rente.de/drv/`.

E65-E68: Obesity and other hyperalimentation

I10-I15: Hypertensive diseases

I20-I25: Ischaemic heart diseases

I60-I69: Cerebrovascular diseases (strokes)

- Diseases of the musculoskeletal system and connective tissue (ICD-10)

  M15-M19: Arthrosis

  M50-M54: Other dorsopathies

- Health care consumption

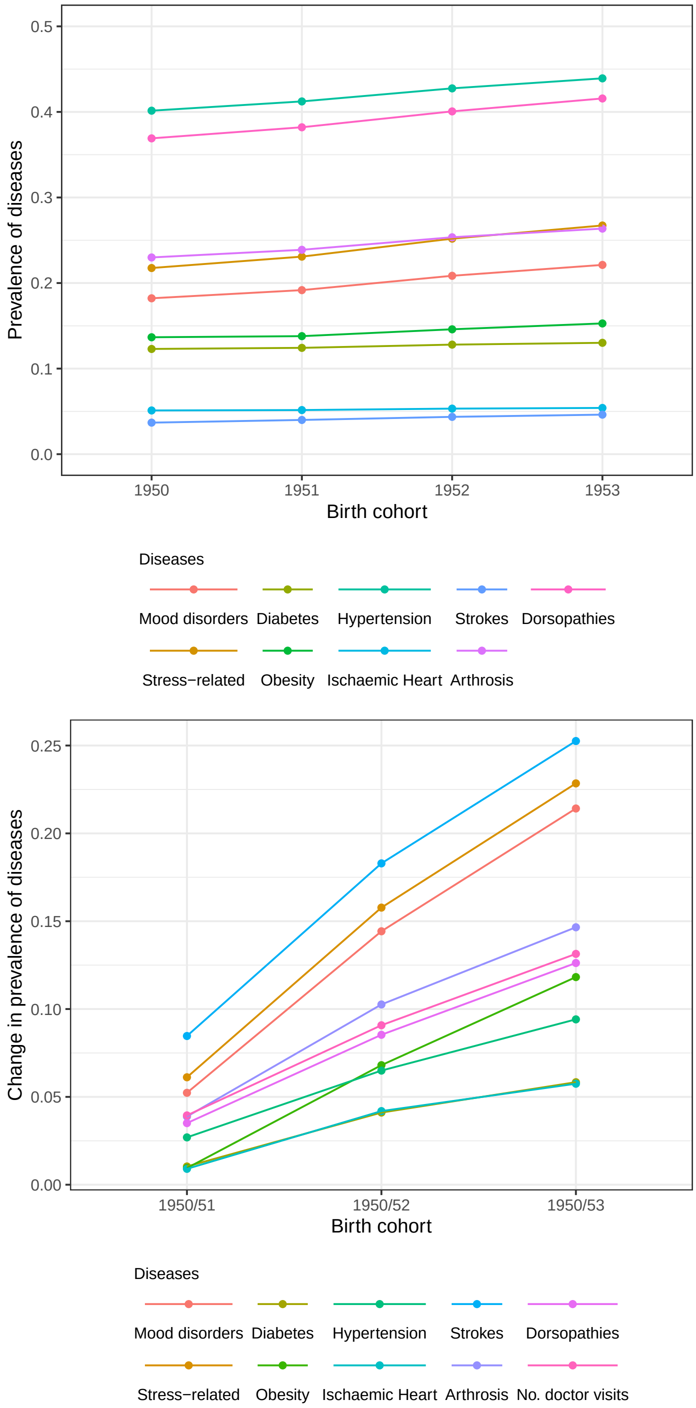
  Doctor visits

Figure 2.1 shows the prevalence of the selected diagnoses within our sample and how they vary across cohorts. The top panel presents the average share of women suffering from a certain disease by birth cohort. The prevalence of diseases in our sample ranges from about 5% (ischaemic heart diseases and strokes) to about 40% (hypertension and other dorsopathies). It is also visible that most diseases have a positive and sizable cohort trend, meaning that younger cohorts have a higher likelihood to be diagnosed with one of the diseases. This pattern becomes clearer in the bottom panel of Figure 2.1, which presents the percentage difference in the prevalence of the diseases compared to cohort 1950. The graphical evidence underlines the importance to control for cohort effects to identify the causal reform effect in the empirical analysis. Additionally, Table 2.A.1 in the Appendix presents sample means for all considered diagnoses for the three different age groups.

2.4 Empirical strategy

In the main analysis we use a DiD approach to estimate the effect of the 1999 pension reform on health outcomes. We complement the analysis using a RDD. The medical literature (e.g., Boland et al., 2015; Doblhammer and Vaupel, 2001) documents that the month of birth is correlated with health outcomes. In the RDD, we can only account for seasonality (month of birth effects) by including quarter of birth as a control variable. This, however, requires an observation period of at least 12 months before and after the cutoff, thus exacerbating the challenge to absorb cohort effects, especially if observations are grouped by month of birth. Therefore, we only present the RDD results in the Appendix and focus in the main analysis on the DiD analysis that explicitly accounts for potential month of birth effects by differencing them out.

Figure 2.1: Prevalence of diagnosed diseases and cohort trends



Notes: The top figure presents the average prevalence of the different diseases among women between age 60 and 62 by birth cohort. The dots represent means. The bottom figure presents the average difference in the prevalence of the diseases and the number of doctor visits compared to birth cohort 1950 for women between age 60 and 62.
Source: KBV, own calculations.

Specifically, as Schönberg and Ludsteck (2014), we define a control group (women born between October 1950 and March 1951) and a treatment group (women born between October 1951 and March 1952). Only women born between January and March are affected by the reform, i.e. they belong to the post reform group. Thus, the interaction between treatment group and being born after the cutoff estimates the effect of the pension reform in the DiD setting. Importantly, the sample only includes individuals born between October 1951 and March 1952 as well as between October 1950 and March 1951, respectively. Thus, birth months between March and October are not included in the sample. This way, we avoid comparing birth months that are rather far away from the reform cutoff in January.

More formally, we estimate the following equation:

$$\begin{aligned} y_{it} =& \alpha^{DiD} + \beta_0^{DiD} Winter_i^{5152} + \beta_1^{DiD} JanFebMar_i + \beta_2^{DiD} Winter_i^{5152} \times JanFebMar_i \\ &+ Z_{it}\delta^{DiD} + \varepsilon_{it}^{DiD} \end{aligned} \tag{2.1}$$

where $Winter_i^{5152}$ indicates whether individual i was born between October 1951 and March 1952. The indicator is zero if individual i was born between October 1950 and March 1951. $JanFebMar_i$ is the reform indicator that is one if individual i was born between January and March and zero otherwise. $Winter_i^{5152} \times JanFebMar_i$ is the interaction between the two indicator variables and turns one for every woman born from January 1952. Thus, the interaction term marks the individuals who are affected by the reform. In addition, we account for age dummies and regional effects[22] captured in Z_{it}.

To test for significance of our results we cluster standard errors by month of birth and we perform multiple hypotheses tests to account for the uncertainty related to the relatively large number of outcome variables. In order to estimate valid causal effects with the difference-and-difference estimator several assumptions need to hold. First, the intervention needs to be unrelated to the outcomes at baseline, which holds in this case by construction as the division into treatment and control group is determined by birthday which is exogenously determined. For the same reason the composition of treatment and control group is stable and there are no spillover effects. Secondly, we provide graphical evidence that the parallel trends assumption holds (parallel trends in the outcomes of treatment and control group prior to the intervention) in the Appendix section 2.A.2.

[22]We include an east/west dummy variable.

2.5 Empirical results

In the following, we present the estimation results of the DiD estimation and discuss how an increase in the retirement age affects the health outcomes defined above. We estimate the effects for different age groups. Our main focus is on the group of 60–62 year old women. Effects are most direct for this group because in younger ages women of neither cohort can enter an old age retirement scheme. Women's health might, however, react to the reform already before reaching the age of 60 because they anticipate and expect to retire only three years later than expected. Therefore, we also study effects at age 59.[23] There are two main channels through which the expectation of retiring only at age 63 could affect health at age 59: First, the effect could be caused by the expectation of working three years longer ("real" retirement effect). Second, cohort 1952 could perceive the reform as unfair as their only slightly older peers can retire three years before them (fairness effect). Thus, effects at age 59 are likely a mixture of both a "real" retirement effect and a fairness effect. In section 2.5.5 we will turn to women aged 63–65. Women born in 1952 or later can enter retirement at age 63, therefore theses results can be interpreted as post employment effects.

In the data, we neither have information on the working history of women nor on their eligibility for the old-age pension for women. Therefore, we identify an intent-to-treat effect (ITT) of the pension reform. According to Geyer and Welteke (2021), about 60 percent of all women born in 1951 were eligible for the old-age pension for women.

2.5.1 Results – Mental health

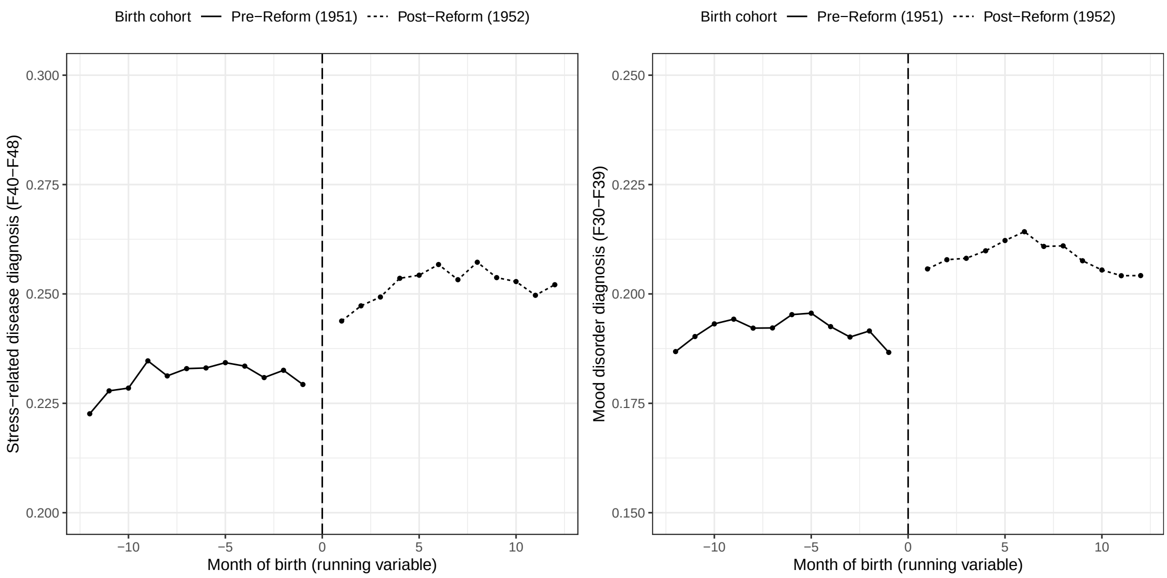
We start with the discussion of the effects of the pension reform on two dimensions of mental health: stress-related mental diseases and mood disorders. The first subsection depicts descriptive, graphical evidence. Thereafter, we present the estimated causal effects of the increase in the ERA based on the DiD. The complementary results of the RDD are presented in Appendix section 2.A.2.

Graphical analysis. Figure 2.2 shows the average share of women aged 60 to 62 who are diagnosed with a stress related or a mood disorder diagnosis by month of birth. For both groups of diseases, there is a distinct and clear jump at the reform cutoff that ranges between one and two percentage points. In addition, there is evidence of seasonality in the trend both before and after the cutoff. This underlines the importance

[23] The anticipation effects could already be present before the age of 59. Unfortunately, we do not have the data to study the effects at younger ages.

of controlling for quarter of birth in addition to potential cohort effects to identify the causal effect of the reform and motivates the DiD approach.

Figure 2.2: Diagnoses of mental and behavioral disorders by month of birth



Notes: The left figure presents the average share of women between age 60 and 62, who got a F40-F48 diagnosis in a given year, for each birth month. The right figure presents the average share of women between age 60 and 62, who got a F30-F39 diagnosis in a given year, for each birth month. The vertical lines represent the cutoff date (01/1952).
Source: KBV, own calculations.

Regression results. The regression results based on the DiD (Table 2.1) confirm the graphical evidence: The effect on stress-related diseases for 60–62 year old women amounts to 0.8 percentage points (3.6 percent relative to the pre-treatment mean). For mood disorder diagnoses, the estimated effect in age group 60–62 are slightly higher (0.9 percentage points), which corresponds to a relative effect of 4.8 percent in relation to the pre-treatment mean. So far, we focus on the effects of the main group of interest, namely 60–62 year old women. As mentioned above, women's health might react to the reform even before reaching the age of 60 because they know that they need to work three years longer. In fact, we find that the effects for 59 year old women are even higher and clearly significant. Thus, anticipation effects are important.

We provide empirical evidence for our identification strategy in Appendix section 2.A.2. First, the pre-reform time trends for the treatment and the control groups for the different diagnoses are very similar (Figure 2.A.1) and, second, the estimates of a placebo test are not significant (Columns 1 and 2 in Table 2.A.2). Specifically, for the placebo test we use the same empirical specification but artificially shift the design by one year and assign the cohort born in the first quarter 1951 as the treatment group after the hypothetical reform.

Table 2.1: DiD: Mental diagnoses

	Stress-related		*Mood disorder*	
	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.008**	0.015***	0.009***	0.014***
	(0.003)	(0.003)	(0.002)	(0.003)
$Winter5152_i$	0.012***	0.006***	0.008***	0.002
	(0.002)	(0.001)	(0.002)	(0.002)
$JanFebMar_i$	0.008***	0.004+	0.009***	0.006*
	(0.002)	(0.002)	(0.002)	(0.002)
Pre-treatment mean	0.222	0.206	0.186	0.17
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for west	yes	yes	yes	yes
Observations	1,738,083	627,391	1,738,083	627,391

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the DiD estimates for women aged 60–62 years and include age dummies and and a West-Germany dummy as control variables. Columns (2) and (4) show the DiD estimates for women at age 59 and include a West-Germany dummy as control variable. All regressions include the cohort indicator, the reform indicator and their interaction term. *Source:* KBV, own calculations.

In Appendix section 2.A.2 we present in addition the results based on the RDD design. The results confirm the findings based on the DiD: The increase in the retirement age has a positive effect on both outcomes, stress related or a mood disorder diagnosis, at age 59 and between ages 60–62 (Table 2.A.4). To further corroborate our findings, we alter the definition of the outcome variables and test, whether noise of erroneous one-time diagnoses or miss-classifications by the medical personnel drive the results. For this exercise, we follow the so-called M2Q criterion and define a person in a calendar year to suffer from a mental disease only if she was diagnosed with such a condition in two quarters of the calendar year. Compared to the main specification, this alternative definition is more conservative because women who were only diagnosed in one quarter in a specific calendar year are not considered to suffer from the condition in this robustness check. Table 2.A.10 in the Appendix shows the results for this exercise using the DiD specification. For both outcomes, the estimated treatment effects at age 59 and age 60–62 are positive and significant, as in the main specification, but slightly smaller.

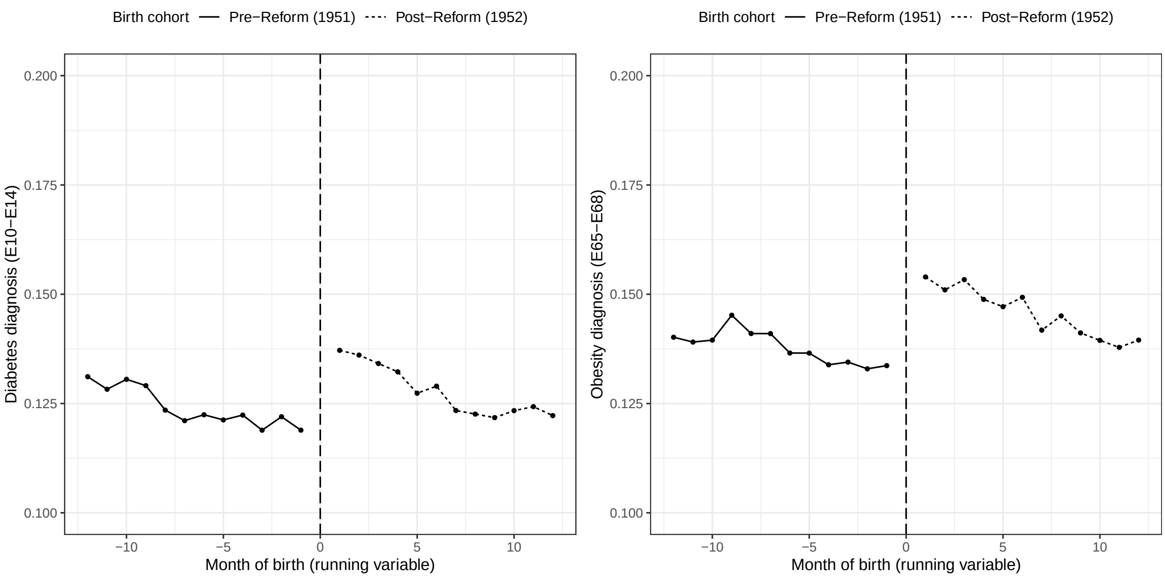
2.5.2 Results – Physical health

In the next step, we analyze the impact on physical health outcomes. We study three groups of physical health outcomes: Nutritional and metabolic diagnoses (diabetes and obesity), musculoskeletal diagnoses (arthrosis and dorsopathies), as well as circulatory and heart diagnoses (hypertension, ischaemic heart diseases and strokes).

Graphical analysis. The graphical analysis reveals the importance of seasonality for the different physical health outcomes and provides mixed evidence about the effect of the 1999 pension reform on physical health. Regarding the nutritional and metabolic outcomes, we observe a strong seasonality pattern (Figure 2.3). Women born early in the year are more likely to be diagnosed with either of the diseases (diabetes and obesity) compared to women born later in the year. This is in line with findings from the medical literature that suggest that environmental reasons, exposure to sunlight, or nutrition are the main drivers for these differences (e.g., Kahn et al., 2009; Phillips and Young, 2000; Vaiserman and Khalangot, 2008; Wattie et al., 2008). Apart from seasonality, there seems to be no clear and strong jump at the reform cutoff.

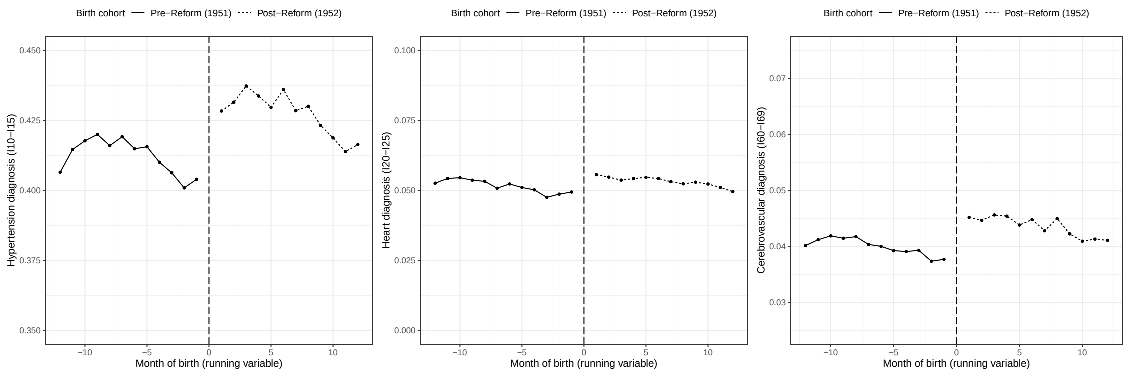
For circulatory and heart diseases the pattern is similar: The graphical evidence does not indicate sizable reform effects (Figure 2.4). In line with Boland et al. (2015), we also find a strong seasonality pattern for hypertension whereas the pattern for heart and cerebrovascular diseases is rather stable. Musculoskeletal diagnoses also show quite strong seasonal fluctuations especially for arthrosis (Figure 2.5). However, there is some evidence of a positive reform effect on both musculoskeletal outcomes under study.

Figure 2.3: Metabolic/nutritional diagnoses by month of birth



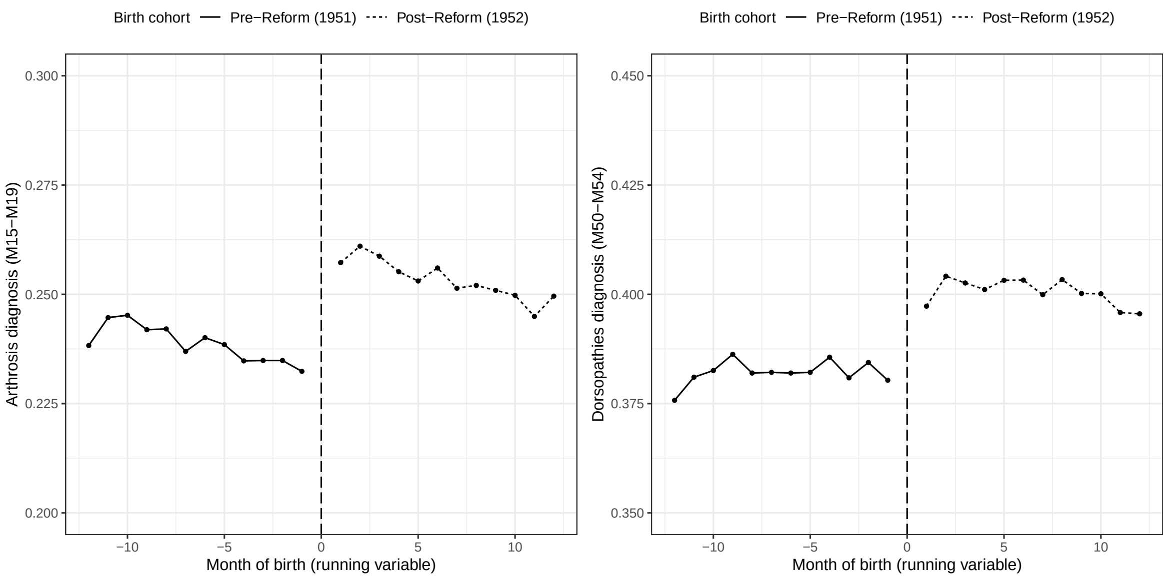
Notes: The left figure presents the average share of women between age 60 and 62, who got a E10-E14 diagnosis in a given year, for each birth month. The right figure presents the average share of women between age 60 and 62, who got a E65-E68 diagnosis in a given year, for each birth month. The vertical lines represent the cutoff date (01/1952).
Source: KBV, own calculations.

Figure 2.4: Circulatory/heart diagnoses by month of birth



Notes: The left figure presents the average share of women between age 60 and 62, who got a I10-I15 diagnosis in a given year, for each birth month. The figure in the middle the average share of women between age 60 and 62, who got a I20-I25 diagnosis in a given year, for each birth month. The right figure presents the average share of women between age 60 and 62, who got a I60-I69 diagnosis in a given year, for each birth month. The vertical lines represent the cutoff date (01/1952).
Source: KBV, own calculations.

Figure 2.5: Musculoskeletal diagnoses by month of birth



Notes: The left figure presents the average share of women between age 60 and 62, who got a M15-M19 diagnosis in a given year, for each birth month. The right figure presents the average share of women between age 60 and 62, who got a M50-M54 diagnosis in a given year, for each birth month. The vertical lines represent the cutoff date (01/1952).
Source: KBV, own calculations.

Regression results. In the following, we present the DiD results. We first cover metabolic and nutritional diseases, then, in the second subsection, we show the effects on circulatory and heart diseases, and the last subsection presents musculoskeletal diseases. Overall, the regression results largely support the insights from the graphical analysis.

Metabolic and nutritional diseases. We find small positive effects of the reform on both health outcomes (Table 2.2). In more detail, for diabetes, the interaction effect in the DiD specification, which captures the effect of the pension reform, is positive and significant (0.3 percentage points for the main effect and 0.5 percentage points for the age-59-effect). Thus, the results suggest that the pension reform has a significant but small effect on the prevalence of diabetes. The pattern is similar for obesity. Again, the point estimates are small and positive but highly significant. The robustness checks which are presented in the Appendix support the identification strategy (pre-reform trends in Figure 2.A.2 and placebo tests in Columns 5 and 6 in Table 4.A.7) and confirm the findings. Specifically, the results for diabetes and obesity are confirmed when using the more conservative definition of the outcome variable (M2Q-criterion in Table 2.A.11).[24]

Table 2.2: DiD: Metabolic/nutritional diagnoses

	Diabetes		*Obesity*	
	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.003*	0.005**	0.010***	0.009***
	(0.002)	(0.002)	(0.001)	(0.001)
$Winter5152_i$	0.002^{+}	0.001	0.003**	-0.002^{*}
	(0.001)	(0.001)	(0.001)	(0.001)
$JanFebMar_i$	0.013***	0.011***	0.010***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Pre-treatment mean	0.123	0.097	0.135	0.123
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for west	yes	yes	yes	yes
Observations	1,738,083	627,391	1,738,083	627,391

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the DiD estimates for women aged 60–62 years and include age dummies and and a West-Germany dummy as control variables. Columns (2) and (4) show the DiD estimates for women at age 59 and include a West-Germany dummy as control variable. All regressions include the cohort indicator, the reform indicator and their interaction term.
Source: KBV, own calculations.

Circulatory and heart diseases. We do not find significant effects on hypertension, ischeamic heart diseases or strokes for 60–62 year old women (Table 2.3). Interestingly,

[24]Note, the point estimates in the RDD (Table 2.A.5) are considerably larger (2 percentage points for the main effect and 1.8 percentage points for the age-59-effect), which is consistent with the strong seasonality pattern presented in the figures above.

the age-59-effects are significant for hypertension (2.4 percentage points) and for strokes (0.2 percentage points). The results for hypertension and heart diseases when using the M2Q-criterion are quite similar to our main specifications. However, for strokes, the estimates turn - at low levels - significant (Table 2.A.12). Overall, we do not find strong evidence that the increase in the retirement age increases the prevalence of the circulatory and heart diseases under study. Further, results for strokes need to be interpreted with caution as we find in the placebo regression (Column 9 in Table 4.A.7) small and positive effects for this outcome which suggests that pre-reform trends of the treatment and the control group are different (see as well Figure 2.A.3). The placebo tests for hypertension and ischeamic heart diseases are not significant.[25]

Table 2.3: DiD: Circulatory/heart diagnoses

	Hypertension		*Heart diagnosis*		*Stroke*	
	Main	Age-59	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.007^{+}	0.024^{***}	0.0004	-0.0003	0.001	0.002^{***}
	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
$Winter5152_i$	0.011^{***}	-0.0005	0.0004	0.0005	0.003^{***}	0.001^{***}
	(0.002)	(0.002)	(0.001)	(0.0004)	(0.001)	(0.0004)
$JanFebMar_i$	0.023^{***}	0.010^{***}	0.006^{***}	0.005^{***}	0.006^{***}	0.004^{***}
	(0.003)	(0.003)	(0.001)	(0.001)	(0.0005)	(0.0004)
Pre-treatment mean	0.403	0.342	0.05	0.04	0.038	0.027
Age group included	60-62 years	59 years	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no	yes	no
Control for west	yes	yes	yes	yes	yes	yes
Observations	1,738,083	627,391	1,738,083	627,391	1,738,083	627,391

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1), (3) and (5) show the DiD estimates for women aged 60–62 years and include age dummies and a West-Germany dummy as control variables. Columns (2), (4) and (6) show the DiD estimates for women at age 59 and include a West-Germany dummy as control variable. All regressions include the cohort indicator, the reform indicator and their interaction term.
Source: KBV, own calculations.

Musculoskeletal diseases. Results for musculoskeletal diseases (arthrosis and dorsopathies) indicate positive effects of the pension reform. We find small and positive effects for both outcomes for women aged 60–62 and for women aged 59 years (Table 2.4). For dorsopathies the anticipation effects are again larger than the main effects. The robustness checks confirm this pattern. First, placebo tests (Column 3–5 in Table 4.A.7) and pre-trends (Figure 2.A.4) support the identification strategy. Second, results using the M2Q-criterion for the definition of the outcome variables are similar

[25] Given the strong seasonal effects presented in Figure 2.4, the positive effects estimated in the RDD (Table 2.A.6) are difficult to interpret.

for arthrosis and dorsopathies (Table 2.A.13). Finally, the results of the RDD (Table 2.A.7) point in the same direction, although the point estimates for arthrosis are slightly larger.

Table 2.4: DiD: Musculoskeletal diagnoses

	Arthrosis		*Dorsopathies*	
	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.008**	0.007*	0.008**	0.021***
	(0.003)	(0.003)	(0.003)	(0.004)
$Winter5152_i$	0.008***	0.004+	0.013***	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
$JanFebMar_i$	0.017***	0.014***	0.012***	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Pre-treatment mean	0.235	0.201	0.374	0.352
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for west	yes	yes	yes	yes
Observations	1,738,083	627,391	1,738,083	627,391

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the DiD estimates for women aged 60–62 years and include age dummies and and a West-Germany dummy as control variables. Columns (2) and (4) show the DiD estimates for women at age 59 and include a West-Germany dummy as control variable. All regressions include the cohort indicator, the reform indicator and their interaction term. *Source:* KBV, own calculations.

2.5.3 Results – Multiple hypothesis testing

Given the relatively large number of health outcomes used in the analysis, we perform multiple-hypothesis-tests using a Bonferroni correction adjustments procedure to the single physical and mental health outcomes. We correct for nine hypotheses (number of diagnoses considered).[26] The multiple hypothesis method confirms our findings of rejecting the null hypothesis for stress-related diseases, mood disorders, obesity, arthrosis and dorsopathies. The results are shown in the Appendix (Table 2.A.14).

2.5.4 Results – Health care consumption

In this section, we turn to the effects of the 1999 pension reform on doctor visits. We measure doctor visits as doctor cases, aggregated at the calendar year level (official term: "Arztfälle"). One doctor case is defined as a treatment of an insured person by a

[26] We choose the Bonferroni correction as our preferred method since this is the most conservative correction procedure. We implement this by using the R-package *p.adjust*.

doctor in a quarter, billed to one public health insurance fund.[27] Thus, if a person visits two different doctors in a quarter, she has two doctor cases in that specific quarter.[28] We 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 the same doctor would have a yearly count of four doctor cases, irrespective of the actual number of visits to this doctor per quarter.[29]

Graphical analysis. Figure 2.6 shows the average number of doctor visits per year for each birth month around the reform cutoff. There is a jump of almost 0.5 doctor visits at the threshold. However, it is important to take into account that the number of doctor visits also varies by about 0.25 doctor visits over birth months on both sides of the discontinuity. Thus, a formal estimation of the causal effect needs to control for month of birth effects and trends.

Regression results. We find a positive and significant effect on healthcare consumption. However, the effect for the main age group is quite small: the number of doctor visits increases by 0.18 visits while the pre-reform mean is 9.43 (Table 2.5). Interestingly, the effect for women aged 59 is more than double the size of the main effect on 60–62 year old women and highly significant. The number of doctor visits increases due to the reform by more than half a doctor visit (Table 2.5). In relative terms, this effect amounts to about 6 percent in relation to the cohort 1951 average of 8.5 visits. Results are again confirmed by the robustness checks presented in Column 10 of Table 4.A.7 (placebo test) and Figure 2.A.5 (pre-trends).

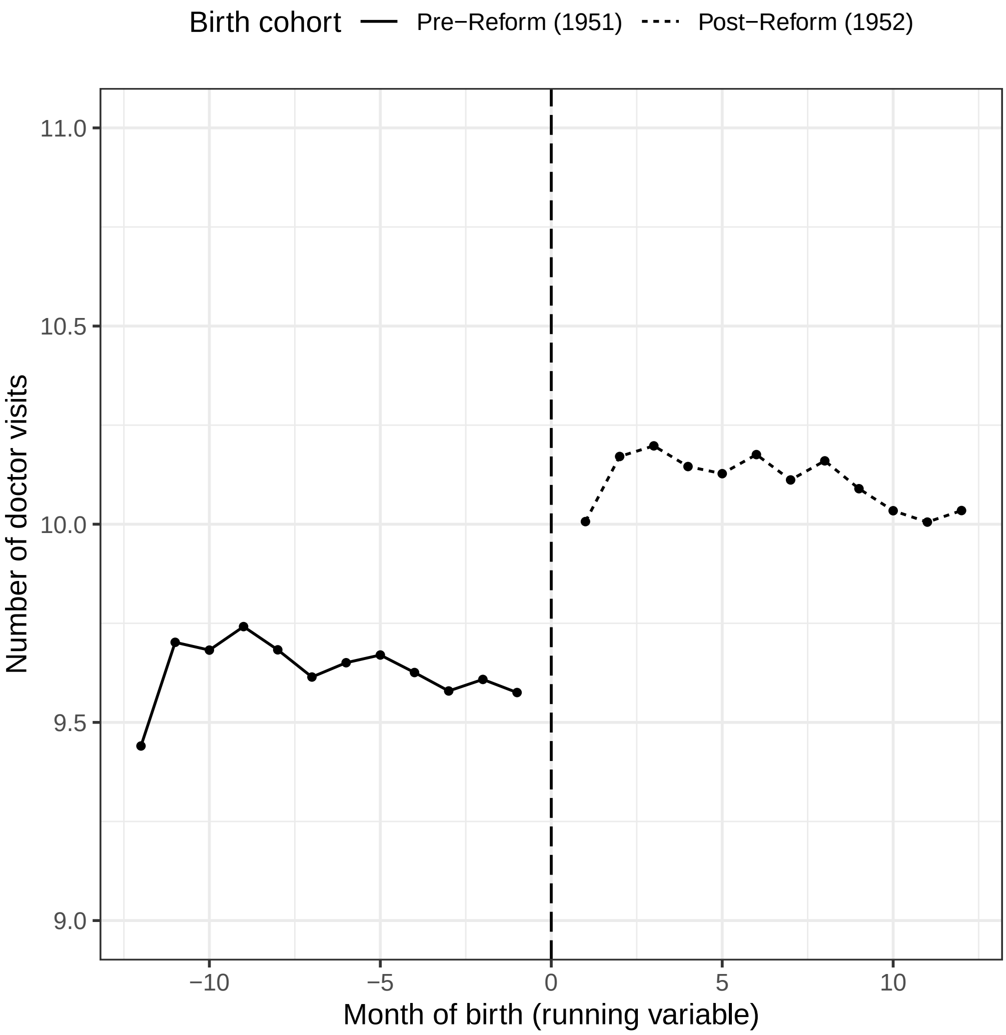
The reasons for the sizable age-59-effect are manifold. One possibility is that women born in 1952 might try to retire early via the disability/invalidity pension schemes in the absence of the old age pension scheme for women. Disability pension is only granted if a person has a reduced earnings capacity and the social-medical assessment is strict. Doctor visits might be indicative of cohort 1952 trying to prove reduced earnings capacity for medical reasons. However, Geyer and Welteke (2021) show that there is no effect of the 1999 pension reform on actual disability pension claims. Thus, despite a possible increase in applications and related doctor visits, the actual claiming behavior is not very different between cohorts 1951 and 1952.

[27]Since doctor cases are recorded this way in the data, we do not have the possibility to define the variable differently for our application.

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

[29]This measure does not capture all doctor visits, thus the observed difference between the two birth cohorts is a lower-bound estimate of the effect of the reform on healthcare consumption.

Figure 2.6: Number of doctor visits by month of birth



Notes: The figure presents the average number of annual doctor visits of women between age 60 and 62 for each birth month. The vertical lines represent the cutoff date (01/1952).
Source: KBV, own calculations.

Another possible reason for differences in healthcare consumption between the cohorts could be different time budgets and time-use decisions in response to the reform. Eligible women born in cohort 1951 know that they can retire at age 60. Thus, they might delay time consuming activities, like (non-urgent) doctor visits from age 59 to their retirement a couple of months later, resulting in fewer doctor visits at age 59. In contrast, women born in 1952 expect to retire only years later, which means that they are less likely to shift time consuming activities from age 59 to age 60. Thus, women born 1952 could have more doctor visits at age 59 than women born in 1951.

Table 2.5: DiD: Number of doctor visits

	Dependent variable: Doctor visits	
	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.180*	0.515***
	(0.087)	(0.086)
$Winter5152_i$	0.328***	0.092*
	(0.026)	(0.039)
$JanFebMar_i$	0.377***	0.161*
	(0.071)	(0.065)
Pre-treatment mean	9.43	8.52
Age group included	60-62 years	59 years
Control for age	yes	no
Control for west	yes	yes
Observations	1,738,083	627,391

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Column (1) shows the DiD estimates for women aged 60–62 years and include age dummies and a West-Germany dummy as control variable. Column (2) shows the DiD estimates for women at age 59 and includes a West-Germany dummy as control variable. All regressions include the cohort indicator, the reform indicator and their interaction term.
Source: KBV, own calculations.

In Table 2.6 we focus in more detail on healthcare consumption and distinguish treatment cases[30] between general practitioners (GP) and specialists. For the main group, women aged 60-62, this analysis shows that the overall positive effect is only related to an increase in the treatment cases for specialists. For women aged 59 years, both specialists and GP treatment cases contribute to the positive effect.

[30]A treatment case is a slightly different measure for healthcare consumption than doctor visits. This measure is available for different specialist groups. One treatment case is defined as a treatment of an insured person by a doctor's office in a quarter, billed to one public health insurance fund.

Table 2.6: DiD: Treatment cases

	Treatment cases		*Treatment cases (GP)*		*Treatment cases (Specialist)*	
	Main	Age-59	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.158*	0.452***	0.012	0.233***	0.146**	0.220***
	(0.077)	(0.077)	(0.026)	(0.025)	(0.053)	(0.055)
$Winter5152_i$	0.273***	0.048	0.111***	−0.031***	0.163***	0.079*
	(0.024)	(0.037)	(0.003)	(0.005)	(0.023)	(0.034)
$JanFebMar_i$	0.309***	0.101+	0.115***	−0.031+	0.194***	0.131**
	(0.063)	(0.059)	(0.019)	(0.018)	(0.045)	(0.042)
Pre-treatment mean	8.501	7.726	2.676	2.398	5.825	5.329
Age group included	60-62 years	59 years	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no	yes	no
Control for west	yes	yes	yes	yes	yes	yes
Observations	1,738,083	627,391	1,738,083	627,391	1,738,083	627,391

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1), (3) and (5) show the DiD estimates for women aged 60–62 years and include age dummies and a West-Germany dummy as control variables. Columns (2), (4) and (6) show the DiD estimates for women at age 59 and include a West-Germany dummy as control variable. All regressions include the cohort indicator, the reform indicator and their interaction term.
Source: KBV, own calculations.

2.5.5 Post-Employment Effects

In the final section we analyze if the effective increase in the retirement age from 60 to 63 has an effect on health outcomes of women aged 63 and older. These results can be interpreted as indirect or medium run effects of the pension reform since at these ages women of both cohorts have access to retirement and are thus not directly affected by the pension reform. Results of the corresponding DiD are presented in Table 2.7, the placebo tests (Table 2.A.3) and the RDD results (Table 2.A.9) are presented in the Appendix.

The results suggest that the increase of the retirement age has a smaller impact on medium run health outcomes of women (Table 2.7). We only find significant effects below the 5% level for mood disorders, arthrosis, dorsopathies and obesity. Recall, for women aged 60-62 years, we have documented significant and robust evidence for an increase in the prevalence of stress-related diseases, mood disorders, dorsopathies, arthrosis and obesity. The effects on mood disorders, dorsopathies, arthrosis and obesity seem to persist also in the medium run. However, effect sizes are smaller (2.4% vs. 4.8% for mood disorders, 2% vs. 3.4% for arthrosis, 1.2% vs. 2.1% for dorsopathies and 4.2% vs. 7.4% for obesity). This pattern suggests that the detrimental health effects of the increase in retirement age are strongest for women directly affected by the pension reform. However, the majority of the effects persist at least until age 65.

Table 2.7: DiD results: 63-65 year olds

	Dependent variable:									
	Stress-related	Mood disorder	Arthrosis	Dorso-pathies	Diabetes	Obesity	Hyper-tension	Heart	Strokes	Doc. visits
$Winter5152_i \times JanFebMar_i$	0.003	0.005**	0.006*	0.005*	0.005+	0.007***	0.005+	−0.001	0.001	0.116+
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.071)
$Winter5152_i$	0.012***	0.009***	0.011***	0.014***	−0.001	0.008***	0.011***	0.001	0.003***	0.328***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.0003)	(0.021)
$JanFebMar_i$	0.009***	0.011***	0.021***	0.014***	0.013***	0.017***	0.024***	0.008***	0.008***	0.433***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.0002)	(0.058)
Pre-treatment mean	0.247	0.208	0.293	0.412	0.166	0.167	0.498	0.07	0.061	10.943
Observations	1,543,601	1,543,601	1,543,601	1,543,601	1,543,601	1,543,601	1,543,601	1,543,601	1,543,601	1,543,601

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. All regressions include the cohort indicator, the reform indicator and their interaction term.
Source: KBV, own calculations.

As the effect sizes decrease with age, our results indicate that the differences in health outcomes between the two cohorts fade out at older ages, i.e. in the long run. A formal analysis of the long run effects remains for future research when data are available.

2.6 Conclusion

This paper provides novel insights about the causal effects of an increasing retirement age on a multi-dimensional and comprehensive set of health outcomes. For the identification, we exploit a large exogenous increase in the ERA for women in Germany. In particular, we focus on the 1999 pension reform that increases the ERA by three years for women born after December 1951.

Previous literature is inconclusive in terms of magnitude and direction of the overall effects of retirement on health. Earlier work often relies on survey data that in general only include subjective and broad health measures. However, health is multi-dimensional and the effects of retirement (reforms) on different health outcomes might, therefore, go into different directions.

Our analyses are based on administrative data from German health insurance funds that include health diagnoses of all publicly insured individuals. We use a sample of women born between 1950 and 1952 who are observed between 2009 and 2018. The data contain all diagnoses in outpatient care during the observation period. Specifically, we identify and consider relevant diagnoses and measures within three dimensions of health outcomes: mental health, physical health, and healthcare consumption.

In the empirical analysis we exploit the variation in the ERA by cohorts in a DiD approach and provide various robustness analyses including placebo tests. The findings reflect the multi-dimensionality of health outcomes but allow for deriving two broader conclusions. We provide evidence that the increase in the retirement age has a negative effect on health outcomes as the prevalence of several diagnoses, e.g., mental health, arthrosis, dorsopathies and obesity, increases. In contrast, we do not find support for an improvement in health related to a prolonged working life since there is no significant evidence of a reduction in the prevalence of any health outcome we consider. These findings are robust to the sensitivity checks, and do not change when correcting for multiple hypothesis testing.

More precisely, we find that the pension reform increased the prevalence of both groups of mental diseases in 60–62 year old women. The effect size amounts to 3.6 percent for stress-related diseases and to 4.8 percent for mood disorders relative to the respective pre-treatment means. The effects for 59 year old women are of similar magnitude and significance. Considering that only about 60% of the women were eligible for the old age pension for women (Geyer and Welteke, 2021), the reform effect on eligible women turns out even larger. For example, scaling the ITT effects with this eligibility rate in a back-of-the-envelope calculation, the effects on stress-related diseases for 60–62 year old women amount to 6 and the effects on mood-disorders to 8 percent.[31]

Within the physical health dimension, our ITT estimates suggest that raising the retirement age increases the prevalence of dorsopathies, arthrosis and obesity at age 60–62 years as well as 59 years. For other physical health outcomes our results are less clear but, as mentioned above, we do not find significant evidence for an improvement in physical health in response to the reform. Furthermore, we find a significant increase in healthcare consumption for 59 year olds following the reform.

Additional analyses on post-employment effects suggest that the effects on mood disorders, dorsopathies, arthrosis and obesity persist also in the medium run. However, effect sizes are smaller for 63–65 year old women compared to 60–62 year old women suggesting that the detrimental health effects do last into retirement but at a lower level.

Increasing the retirement age is controversially discussed in politics and society. Our results inform this debate, as health implications are an important aspect. For future pension reforms, policy makers should keep in mind that a prolonged working life might have considerable negative health consequences, particularly for mental health. Further research is needed to empirically identify the mechanisms behind our findings. One important mechanism is certainly related to the prolonged duration in the labor market. This effect operates through different channels which we cannot differentiate with the data at hand. The majority of treated women stays longer in employment which might affect health. However, the prolonged status in unemployment could as well impact health. Moreover, the sizable effects for several outcomes before the retirement age suggest that expectation effects are important. These expectation effects are in line with previous literature, e.g. Grip et al. (2012) find that a change in the retirement system in the Netherlands affecting 62 years olds already led to increases in the depression rates among 58/59 year olds.[32] Policies need to take this into consideration. Targeted health programs that support different groups in the labor market in dealing

[31]Note, for this back-of-the-envelope calculation we have assumed that eligible and non-eligible women are comparable. Given that by definition the non-eligible women have a shorter employment history this assumption is likely not too hold. Therefore, these calculations need to be interpreted as approximations.

[32]In contrast, Bauer and Eichenberger (2021) document negative pre-retirement health effects following a reform that lowered the retirement age from 65 to 60 for construction workers in Switzerland.

with stress or providing sport and exercise programs could counteract the negative effects. Another solution might be to extend old-age-part-time work to smooth the transition into retirement. However, in addition to the measures close to retirement it is important to target individuals already earlier in the life cycle and to provide opportunities to invest into human capital and health. This would allow individuals to prepare for a longer working life.

In future research, it would be important to assess whether these multi-dimensional health effects further differ by socioeconomic characteristics. The literature shows that such characteristics may matter for the health effects of retirement (see e.g., Etgeton and Hammerschmid, 2019, and references therein). The data we use only includes very limited individual characteristics beyond health. Thus, with the data at hand, assessing the socioeconomic gradient and potential mechanisms is not possible. Furthermore, it would be interesting to analyze the effects at ages older that 65 years to understand how persistent the effects are.

2.A Appendix

2.A.1 Descriptive results

Table 2.A.1: Outcomes

	59 years	60-62 years	63-65 years
Mental diagnoses			
Stress-related diseases	0.22 (0.41)	0.23 (0.42)	0.26 (0.44)
Mood disorders	0.18 (0.38)	0.19 (0.40)	0.22 (0.41)
Metabolic/nutritional diagnoses			
Diabetes	0.10 (0.30)	0.13 (0.33)	0.17 (0.37)
Obesity	0.13 (0.33)	0.14 (0.35)	0.17 (0.38)
Circulatory/heart diagnoses			
Hypertension	0.35 (0.48)	0.41 (0.49)	0.51 (0.50)
Heart	0.04 (0.20)	0.05 (0.22)	0.07 (0.26)
Strokes	0.03 (0.17)	0.04 (0.20)	0.06 (0.24)
Musculoskeletal diagnoses			
Arthrosis	0.20 (0.40)	0.24 (0.43)	0.30 (0.46)
Dorsopathies	0.36 (0.48)	0.38 (0.49)	0.42 (0.49)
Healthcare consumption			
Doctor visits	8.75 (8.45)	9.66 (8.60)	11.72 (8.72)
Treatment cases	7.91 (7.50)	8.66 (7.52)	9.90 (7.48)
Treatment cases (GP)	2.47 (2.19)	2.72 (2.18)	3.11 (2.08)
Treatment cases (Specialist)	5.44 (5.99)	5.94 (6.09)	6.79 (6.27)
Observations	1,885,051	5,221,811	4,637,760

Notes: Reported are means and standard deviations in parentheses. Treatment cases are defined as "A treatment case is a slightly different measure for healthcare consumption than doctor visits. This measure is available for different specialist groups. One treatment case is defined as a treatment of an insured person by a doctor's office in a quarter, billed to one public health insurance fund." The means include birth cohorts 1950-1952.

Source: KBV, own calculations.

2.A.2 Robustness

Table 2.A.2: DiD Placebo: 60-62 year olds

	Dependent variable:									
	Stress-related	Mood disorder	Arthrosis	Dorso-pathies	Diabetes	Obesity	Hyper-tension	Heart	Strokes	Doc. visits
$Winter5152_i \times JanFebMar_i$	0.001	-0.003	0.001	0.003	-0.004	0.001	0.003	0.0001	0.002^{*}	0.126
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)	(0.005)	(0.001)	(0.001)	(0.128)
$Winter5152_i$	0.013^{***}	0.013^{***}	0.008^{***}	0.011^{***}	0.003^{**}	-0.003^{*}	0.006^{**}	-0.0001	0.001^{*}	0.276^{***}
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.032)
$JanFebMar_i$	0.007^{*}	0.012^{***}	0.016^{***}	0.009^{*}	0.017^{***}	0.009^{***}	0.020^{***}	0.006^{***}	0.004^{***}	0.252^{*}
	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)	(0.004)	(0.0003)	(0.001)	(0.106)
Pre-treatment mean	0.209	0.175	0.226	0.362	0.123	0.138	0.395	0.051	0.036	9.084
Observations	1,779,704	1,779,704	1,779,704	1,779,704	1,779,704	1,779,704	1,779,704	1,779,704	1,779,704	1,779,704

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. All regressions include the cohort indicator, the reform indicator and their interaction term.

Source: KBV, own calculations.

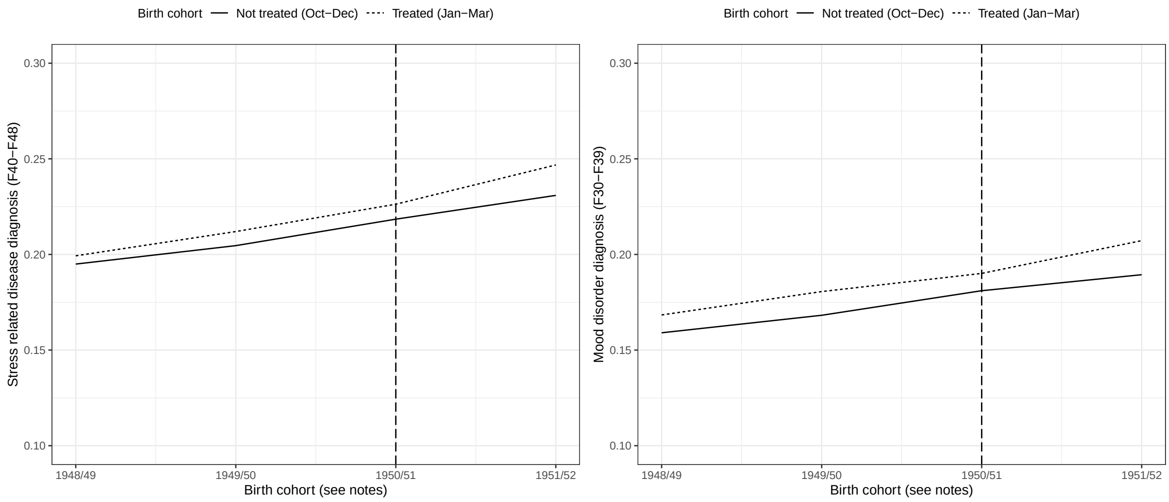
Table 2.A.3: DiD Placebo: 63-65 year olds

	Dependent variable:									
	Stress-related	Mood disorder	Arthrosis	Dorso-pathies	Diabetes	Obesity	Hyper-tension	Heart	Strokes	Doc. visits
$Winter5152_i \times JanFebMar_i$	0.001	-0.004^{*}	-0.001	-0.002	-0.006	-0.001	-0.001	-0.0001	0.001	0.052
	(0.004)	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.004)	(0.001)	(0.001)	(0.107)
$Winter5152_i$	0.014^{***}	0.015^{***}	0.011^{***}	0.016^{***}	0.004^{***}	0.005^{***}	0.010^{***}	0.0004	0.003^{***}	0.360^{***}
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.028)
$JanFebMar_i$	0.008^{**}	0.016^{***}	0.022^{***}	0.016^{***}	0.019^{***}	0.018^{***}	0.025^{***}	0.008^{***}	0.006^{***}	0.383^{***}
	(0.003)	(0.001)	(0.001)	(0.003)	(0.003)	(0.001)	(0.003)	(0.0004)	(0.001)	(0.090)
Pre-treatment mean	0.232	0.196	0.282	0.397	0.165	0.162	0.489	0.07	0.057	10.554
Observations	1,578,369	1,578,369	1,578,369	1,578,369	1,578,369	1,578,369	1,578,369	1,578,369	1,578,369	1,578,369

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. All regressions include the cohort indicator, the reform indicator and their interaction term.

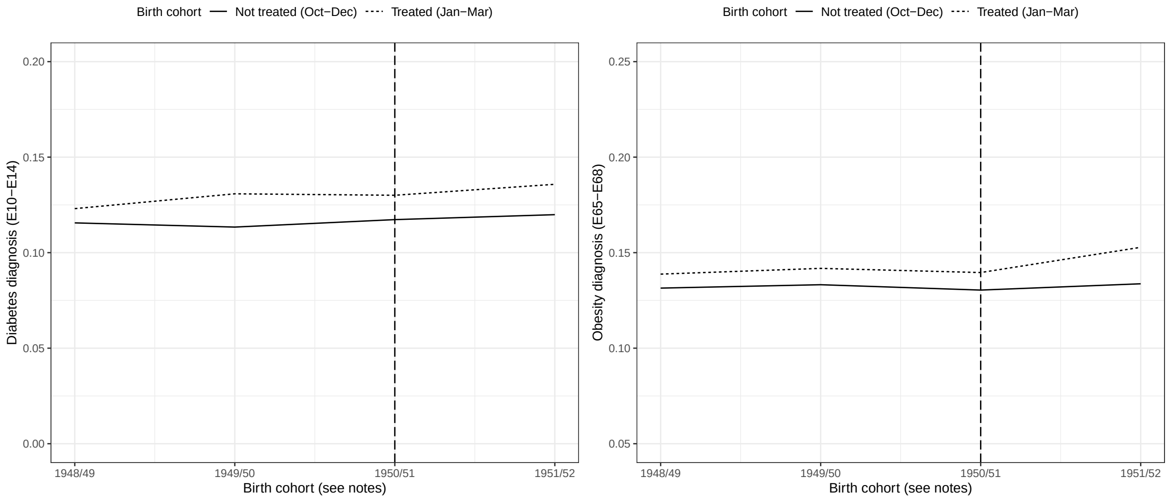
Source: KBV, own calculations.

Figure 2.A.1: Diagnoses of mental and behavioral disorders in treatment and control group



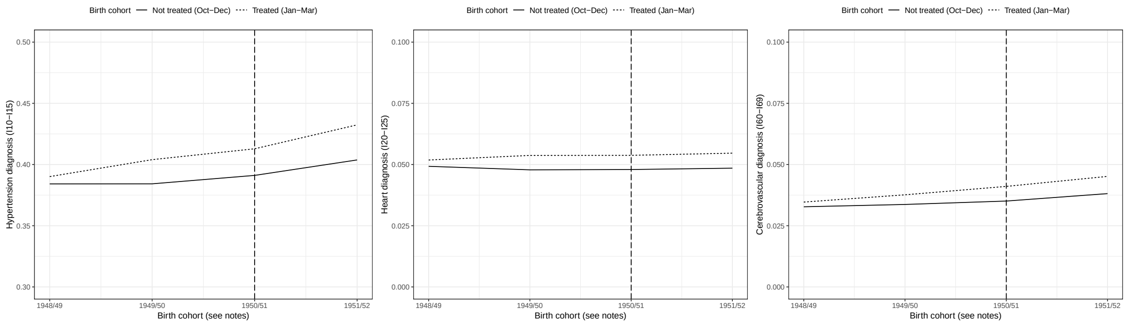
Notes: The left figure presents the average share of women between age 60 and 62, who got a F40-F48 diagnosis in a given year, for each birth cohort. The right figure presents the average share of women between age 60 and 62, who got a F30-F39 diagnosis in a given year, for each birth cohort. The vertical lines represent the cutoff date (01/1952). Birth cohort 1948/49 represents women born between October to December 1948 (control group) and January and March 1949 (treatment group). Accordingly, birth cohorts 1949/50 represent women born between October to December 1949 and January and March 1950, birth cohorts 1949/50 represent women born between October to December 1950 and January and March 1951 and birth cohorts 1951/52 represent women born between October to December 1951 and January and March 1952.
Source: KBV, own calculations.

Figure 2.A.2: Metabolic/nutritional diagnoses in treatment and control group



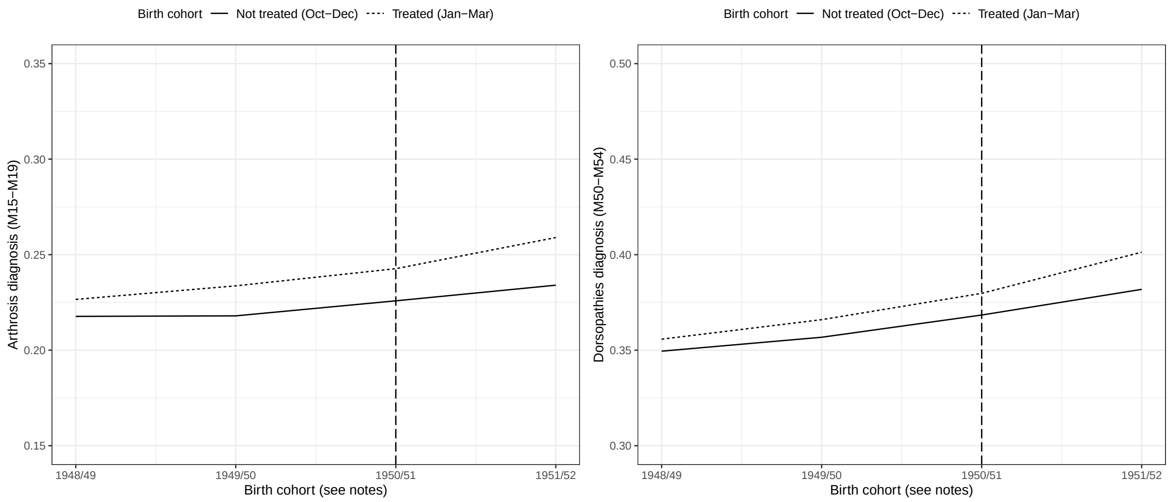
Notes: The left figure presents the average share of women between age 60 and 62, who got a E10-E14 diagnosis in a given year, for each birth cohort. The right figure presents the average share of women between age 60 and 62, who got a E65-E68 diagnosis in a given year, for each birth cohort. The vertical lines represent the cutoff date (01/1952). Birth cohort 1948/49 represents women born between October to December 1948 (control group) and January and March 1949 (treatment group). Accordingly, birth cohorts 1949/50 represent women born between October to December 1949 and January and March 1950, birth cohorts 1949/50 represent women born between October to December 1950 and January and March 1951 and birth cohorts 1951/52 represent women born between October to December 1951 and January and March 1952.
Source: KBV, own calculations.

Figure 2.A.3: Circulatory/heart diagnoses in treatment and control group



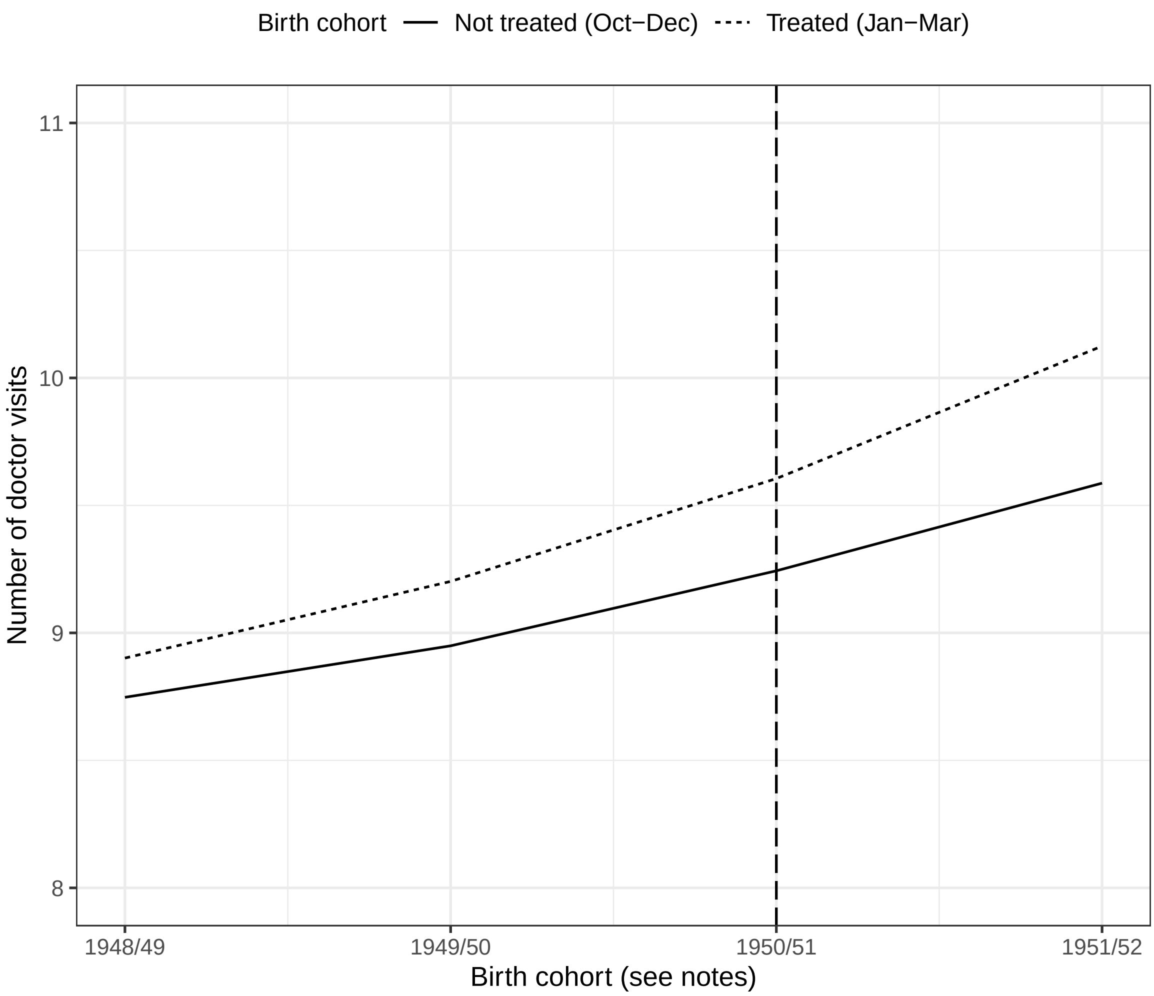
Notes: The left figure presents the average share of women between age 60 and 62, who got a I10-I15 diagnosis in a given year, for each birth cohort. The figure in the middle the average share of women between age 60 and 62, who got a I20-I25 diagnosis in a given year, for each birth cohort. The right figure presents the average share of women between age 60 and 62, who got a I60-I69 diagnosis in a given year, for each birth cohort. The vertical lines represent the cutoff date (01/1952). Birth cohort 1948/49 represents women born between October to December 1948 (control group) and January and March 1949 (treatment group). Accordingly, birth cohorts 1949/50 represent women born between October to December 1949 and January and March 1950, birth cohorts 1949/50 represent women born between October to December 1950 and January and March 1951 and birth cohorts 1951/52 represent women born between October to December 1951 and January and March 1952.
Source: KBV, own calculations.

Figure 2.A.4: Musculoskeletal diagnoses in treatment and control group



Notes: The left figure presents the average share of women between age 60 and 62, who got a M15-M19 diagnosis in a given year, for each birth cohort. The right figure presents the average share of women between age 60 and 62, who got a M50-M54 diagnosis in a given year, for each birth cohort. The vertical lines represent the cutoff date (01/1952). Birth cohort 1948/49 represents women born between October to December 1948 (control group) and January and March 1949 (treatment group). Accordingly, birth cohorts 1949/50 represent women born between October to December 1949 and January and March 1950, birth cohorts 1949/50 represent women born between October to December 1950 and January and March 1951 and birth cohorts 1951/52 represent women born between October to December 1951 and January and March 1952.
Source: KBV, own calculations.

Figure 2.A.5: Number of doctor visits in treatment and control group



Notes: The figure presents the average number of annual doctor visits of women between age 60 and 62 for each birth cohort. The vertical lines represent the cutoff date (01/1952). Birth cohort 1948/49 represents women born between October to December 1948 (control group) and January and March 1949 (treatment group). Accordingly, birth cohorts 1949/50 represent women born between October to December 1949 and January and March 1950, birth cohorts 1949/50 represent women born between October to December 1950 and January and March 1951 and birth cohorts 1951/52 represent women born between October to December 1951 and January and March 1952.
Source: KBV, own calculations.

Regression discontinuity approach

In this Appendix we present the results of the RDD. As discussed in the main text, within the RDD it is difficult to account for seasonality effects. Still, for completeness it is informative to consider the results of the RDD. We implement the RDD according to the following equation:

$$y_{it} = \alpha^{RDD} + \beta^{RDD} D_i + \gamma_0^{RDD} f(M_i - c) + \gamma_1^{RDD} D_i f(M_i - c) + X_{it}\delta^{RDD} + \varepsilon_{it}^{RDD} \quad (2.2)$$

D_i is a dummy specifying treatment that is equal to 1 if a woman is born in January 1952 or later, and 0 otherwise. A woman's month of birth is described by M_i and c is the cutoff date for the increase in early retirement age (ERA, January 1952). The function f represents the trend in the running variable. In our main specification, we include a linear and quadratic cohort trend. This function is interacted with the treatment variable D_i to allow for different slopes before and after the cutoff. In addition, we account for further explanatory variables (X), including quarter of birth and age. The outcome variable y_{it} is defined as an indicator variable that is equal to one if the disease of interest was diagnosed at least once during a calendar year.

Table 2.A.4: RDD-results: Mental diagnoses

	Stress-related		*Mood disorder*	
	Main	Age-59	Main	Age-59
D_i	0.012***	0.014***	0.016***	0.014***
	(0.004)	(0.004)	(0.005)	(0.004)
$Birthmonths$	−0.001*	−0.002***	−0.002***	−0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
$(Birthmonths)^2$	−0.0001***	−0.0002***	−0.0002***	−0.0003***
	(0.00004)	(0.00004)	(0.00005)	(0.0001)
$D_i \times (Birthmonths)$	0.004***	0.005***	0.005***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
$D_i \times (Birthmonths)^2$	−0.00005	0.00003	−0.00002	−0.00000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Pre-treatment mean	0.231	0.212	0.192	0.173
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for birth season	yes	yes	yes	yes
Control for west	yes	yes	yes	yes
Observations	3,429,155	1,235,612	3,429,155	1,235,612

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the RDD estimates for women aged 60–62 years and include age dummies, birth quarter dummies and a West-Germany dummy as control variables. Columns (2) and (4) show the RDD estimates for women at age 59 and include birth quarter dummies and a West-Germany dummy as control variables as control variables. All regressions include linear and quadratic cohort trends in the running variable on both sides of the policy cut-off.
Source: KBV, own calculations.

Table 2.A.5: RDD-results: Metabolic/nutritional diagnoses

	Diabetes		*Obesity*	
	Main	Age-59	Main	Age-59
D_i	0.021***	0.019***	0.020***	0.018***
	(0.005)	(0.005)	(0.005)	(0.003)
Birthmonths	−0.0004	−0.0004	−0.001	−0.001+
	(0.001)	(0.001)	(0.001)	(0.001)
$(Birthmonths)^2$	0.0001	0.00004	−0.00003	−0.00003
	(0.00005)	(0.00005)	(0.00005)	(0.00004)
$D_i \times (Birthmonths)$	−0.002*	−0.002	−0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
$D_i \times (Birthmonths)^2$	0.0001	0.00002	0.0001*	0.0001*
	(0.00005)	(0.00005)	(0.00004)	(0.00004)
Pre-treatment mean	0.124	0.098	0.138	0.125
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for birth season	yes	yes	yes	yes
Control for west	yes	yes	yes	yes
Observations	3,429,155	1,235,612	3,429,155	1,235,612

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the RDD estimates for women aged 60–62 years and include age dummies, birth quarter dummies and a West-Germany dummy as control variables. Columns (2) and (4) show the RDD estimates for women at age 59 and include birth quarter dummies and a West-Germany dummy as control variables as control variables. All regressions include linear and quadratic cohort trends in the running variable on both sides of the policy cut-off.
Source: KBV, own calculations.

Table 2.A.6: RDD-results: Circulatory/heart diagnoses

	Hypertension		*Heart diagnosis*		*Stroke*	
	Main	Age-59	Main	Age-59	Main	Age-59
D_i	0.016*	0.020**	0.007*	0.008**	0.006***	0.006***
	(0.008)	(0.008)	(0.003)	(0.003)	(0.002)	(0.002)
Birthmonths	−0.004**	−0.004**	−0.001	−0.001+	−0.001*	−0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0003)	(0.0003)
$(Birthmonths)^2$	−0.0004***	−0.0003***	−0.00000	−0.00004	−0.0001*	−0.0001*
	(0.0001)	(0.0001)	(0.00005)	(0.00004)	(0.00002)	(0.00002)
$D_i \times (Birthmonths)$	0.008***	0.008***	0.0003	0.001	0.001*	0.001*
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$D_i \times (Birthmonths)^2$	0.0001	0.00003	−0.00001	0.00001	0.00001	0.00001
	(0.0001)	(0.0001)	(0.00003)	(0.00003)	(0.00002)	(0.00002)
Pre-treatment mean	0.412	0.347	0.052	0.041	0.04	0.028
Age group included	60-62 years	59 years	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no	yes	no
Control for birth season	yes	yes	yes	yes	yes	yes
Control for west	yes	yes	yes	yes	yes	yes
Observations	3,429,155	1,235,612	3,429,155	1,235,612	3,429,155	3,429,155

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1), (3) and (5) show the RDD estimates for women aged 60–62 years and include age dummies, birth quarter dummies and a West-Germany dummy as control variables. Columns (2), (4) and (6) show the RDD estimates for women at age 59 and include birth quarter dummies and a West-Germany dummy as control variables. All regressions include linear and quadratic cohort trends in the running variable on both sides of the policy cut-off.

Source: KBV, own calculations.

Table 2.A.7: RDD-results: Musculoskeletal diagnoses

	Arthrosis		*Dorsopathies*	
	Main	Age-59	Main	Age-59
D_i	0.022***	0.022***	0.012*	0.012*
	(0.005)	(0.005)	(0.006)	(0.006)
Birthmonths	−0.003***	−0.003***	-0.001^{+}	-0.001^{+}
	(0.001)	(0.001)	(0.001)	(0.001)
$(Birthmonths)^2$	−0.0002**	−0.0002**	−0.0002**	−0.0002**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$D_i \times (Birthmonths)$	0.004**	0.004**	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
$D_i \times (Birthmonths)^2$	0.0001	0.0001	−0.0001	−0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Pre-treatment mean	0.239	0.203	0.382	0.354
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for birth season	yes	yes	yes	yes
Control for west	yes	yes	yes	yes
Observations	3,429,155	3,429,155	3,429,155	3,429,155

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the RDD estimates for women aged 60–62 years and include age dummies, birth quarter dummies and a West-Germany dummy as control variables. Columns (2) and (4) show the RDD estimates for women at age 59 and include birth quarter dummies and a West-Germany dummy as control variables as control variables. All regressions include linear and quadratic cohort trends in the running variable on both sides of the policy cut-off.
Source: KBV, own calculations.

Table 2.A.8: Control for west RDD-results: Number of doctor visits

	Dependent variable: Doctor visits	
	Main	Age-59
D_i	0.297*	0.610***
	(0.151)	(0.151)
Birthmonths	−0.054*	−0.085***
	(0.024)	(0.024)
$(Birthmonths)^2$	−0.005*	−0.007**
	(0.002)	(0.002)
$D_i \times (Birthmonths)$	0.135**	0.155***
	(0.043)	(0.043)
$D_i \times (Birthmonths)^2$	0.00005	0.001
	(0.002)	(0.002)
Pre-treatment mean	9.631	8.606
Age group included	60-62 years	59 years
Control for age	yes	no
Control for birth season	yes	yes
Control for west	yes	yes
Observations	3,429,155	1,235,612

Notes: $^+$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Column (1) shows the RDD estimates for women aged 60–62 years and include age dummies, birth quarter dummies and a West-Germany dummy as control variables. Column (2) shows the RDD estimates for women at age 59 and include birth quarter dummies and a West-Germany dummy as control variables. Both regressions include linear and quadratic cohort trends in the running variable on both sides of the policy cut-off.
Source: KBV, own calculations.

Table 2.A.9: RDD-results: 63-65 year olds

	Dependent variable:									
	Stress-related	Mood disorder	Arthrosis	Dorso-pathies	Diabetes	Obesity	Hyper-tension	Heart	Strokes	Doc. visits
D_i	0.009*	0.013*	0.029***	0.018**	0.026***	0.027***	0.022**	0.009*	0.005**	0.382**
	(0.004)	(0.005)	(0.007)	(0.006)	(0.007)	(0.005)	(0.007)	(0.004)	(0.002)	(0.134)
$Birthmonths$	−0.003***	−0.002+	−0.003**	−0.003***	0.00000	−0.001+	−0.004**	−0.001	−0.001*	−0.058*
	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0003)	(0.024)
$(Birthmonths)^2$	−0.0002***	−0.0001*	−0.0002+	−0.0002***	0.0001	−0.0001	−0.0003***	−0.00001	−0.00003+	−0.005*
	(0.00005)	(0.0001)	(0.0001)	(0.00004)	(0.0001)	(0.00004)	(0.0001)	(0.0001)	(0.00002)	(0.002)
$D_i \times (Birthmonths)$	0.007***	0.005***	0.004*	0.006***	−0.005**	−0.001	0.008***	0.001	0.001***	0.126**
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.0004)	(0.039)
$D_i \times (Birthmonths)^2$	−0.0001	−0.0001+	0.00004	−0.00005	0.0001	0.0001*	−0.00002	−0.00000	−0.00004	−0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)	(0.00003)	(0.002)
Pre-treatment mean	0.255	0.214	0.298	0.421	0.164	0.172	0.507	0.071	0.063	11.145
Age group included	63-65 years	63-65 years	63-65 years	63-65 years	63-65 years	63-65 years	63-65 years	63-65 years	63-65 years	63-65 years
Control for age	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Control for birth season	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Control for west	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	3,047,412	3,047,412	3,047,412	3,047,412	3,047,412	3,047,412	3,047,412	3,047,412	3,047,412	3,047,412

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. All regressions include linear and quadratic cohort trends in the running variable on both sides of the policy cut-off and age dummies, birth quarter and a West-Germany dummy as control variables.
Source: KBV, own calculations.

Table 2.A.10: M2Q DiD-results: Mental diagnoses

	Stress-related disease		*Mood disorder*	
	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.007**	0.006**	0.007**	0.010***
	(0.002)	(0.002)	(0.002)	(0.003)
$Winter5152_i$	0.010***	0.007***	0.008***	0.003+
	(0.002)	(0.001)	(0.001)	(0.002)
$JanFebMar_i$	0.008***	0.008***	0.010***	0.007***
	(0.002)	(0.001)	(0.001)	(0.002)
Pre-treatment mean	0.143	0.128	0.144	0.125
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for west	yes	yes	yes	yes
Observations	1,738,083	627,391	1,738,083	627,391

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the DiD estimates for women aged 60–62 years and include age dummies and a West-Germany dummy as control variables. Columns (2) and (4) show the DiD estimates for women at age 59. All regressions include the cohort indicator, the reform indicator and their interaction term. The outcome variables are defined according to the M2Q-criterion.

Source: KBV, own calculations.

Table 2.A.11: M2Q DiD-results: Metabolic/nutritional diagnoses

	Diabetes		*Obesity*	
	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.003^{*}	0.005^{**}	0.008^{***}	0.006^{***}
	(0.002)	(0.002)	(0.001)	(0.001)
$Winter5152_i$	0.003^{*}	0.002	0.003^{***}	-0.0004
	(0.001)	(0.001)	(0.001)	(0.001)
$JanFebMar_i$	0.012^{***}	0.010^{***}	0.008^{***}	0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Pre-treatment mean	0.111	0.111	0.097	0.097
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for west	yes	yes	yes	yes
Observations	1,738,083	627,391	1,738,083	627,391

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the DiD estimates for women aged 60–62 years and include age dummies and a West-Germany dummy as control variables. Columns (2) and (4) show the DiD estimates for women at age 59. All regressions include the cohort indicator, the reform indicator and their interaction term. The outcome variables are defined according to the M2Q-criterion.

Source: KBV, own calculations.

Table 2.A.12: M2Q DiD-results: Circulatory/heart diagnoses

	Hypertension		*Heart diagnosis*		*Stroke*	
	Main	Age-59	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.006^{+}	0.025^{***}	0.0004	0.001	0.001^{*}	0.002^{***}
	(0.004)	(0.004)	(0.001)	(0.001)	(0.0004)	(0.001)
$Winter5152_i$	0.012^{***}	0.0004	0.0004	0.0005	0.002^{***}	0.001^{+}
	(0.002)	(0.003)	(0.001)	(0.0004)	(0.0003)	(0.0004)
$JanFebMar_i$	0.023^{***}	0.011^{***}	0.005^{***}	0.003^{***}	0.004^{***}	0.002^{***}
	(0.003)	(0.003)	(0.001)	(0.001)	(0.0003)	(0.0003)
Pre-treatment mean	0.357	0.296	0.037	0.028	0.025	0.017
Age group included	60-62 years	59 years	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no	yes	no
Observations	1,738,083	627,391	1,738,083	627,391	1,738,083	627,391

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1), (3) and (5) show the DiD estimates for women aged 60–62 years and include age dummies and a West-Germany dummy as control variables. Columns (2), (4) and (6) show the DiD estimates for women at age 59. All regressions include the cohort indicator, the reform indicator and their interaction term. The outcome variables are defined according to the M2Q-criterion.
Source: KBV, own calculations.

Table 2.A.13: M2Q DiD-results: Musculoskeletal diagnoses

	Arthrosis		*Dorsopathies*	
	Main	Age-59	Main	Age-59
$Winter5152_i \times JanFebMar_i$	0.007^{***}	0.005^{*}	0.009^{***}	0.013^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
$Winter5152_i$	0.006^{***}	0.003	0.009^{***}	0.002
	(0.001)	(0.002)	(0.001)	(0.002)
$JanFebMar_i$	0.015^{***}	0.011^{***}	0.012^{***}	0.006^{***}
	(0.002)	(0.002)	(0.001)	(0.001)
Pre-treatment mean	0.166	0.137	0.261	0.24
Age group included	60-62 years	59 years	60-62 years	59 years
Control for age	yes	no	yes	no
Control for west	yes	yes	yes	yes
Observations	1,738,083	627,391	1,738,083	627,391

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Columns (1) and (3) show the DiD estimates for women aged 60–62 years and include age dummies and West-Germany as control variables. Columns (2) and (4) show the DiD estimates for women at age 59. All regressions include the cohort indicator, the reform indicator and their interaction term. The outcome variables are defined according to the M2Q-criterion.
Source: KBV, own calculations.

Table 2.A.14: Bonferroni correction for multiple hypothesis testing in DiD - P-values

	60-62 years		*59 years*	
	without correction	Bonferroni	without correction	Bonferroni
Stress-related diseases	0.0042**	0.0376*	0.0000***	0.0000***
Mood disorder	0.0003***	0.0031**	0.0000***	0.0003***
Diabetes	0.0470*	0.4227	0.0075**	0.0672+
Obesity	0.0000***	0.0000***	0.0000***	0.0000***
Hypertension	0.0567+	0.51106	0.0000***	0.0000***
Ischaemic heart diseases	0.6744	1.0000	0.7264	1.000
Stroke	0.1551	1.0000	0.0007***	0.007**
Arthrosis	0.0016**	0.0140*	0.0203*	0.1830
Other dorsopathies	0.0033**	0.0300*	0.0000***	0.0000***

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Columns (1) and (3) show the p-values retrieved from the baseline DiD estimation. The underlying standard errors are clustered on month of birth (running variable) and robust. Columns (2) and (4) show the Bonferroni-corrected p-values.
Source: KBV, own calculations.

CHAPTER 3

Does Grandparenting Pay off for the Next Generations? Intergenerational Effects of Grandparental Care[1]

3.1 Introduction

In light of the increase in longevity, parents today are more likely than in the past to live for many years while their children are adults and parents themselves. Thus, Western societies are experiencing an increase in grandparent-grandchild exposure (e.g., Lowenstein and Bengtson, 2003; Song and Mare, 2019). As a result, today's grandparents are in a better position than previous generations to play an important role in the lives of their children and grandchildren (e.g., Chapman et al., 2018). While grandparents are the most important source of emotional and material support for adult children, they also often represent the most affordable and flexible source of informal childcare for their grandchildren (e.g., Fergusson et al., 2008).

In many OECD countries, grandparents act as the third largest caregiver after parental care and daycare (OECD, 2019a).[2] This is the case in the US but also in continental European countries (Hank and Buber, 2009), although there are significant

[1]This chapter is joint work with C. Katharina Spiess (BiB Wiesbaden, Johannes Gutenberg University, Mainz) and Elena Ziege (BiB Wiesbaden, Johannes Gutenberg University, Mainz).We are grateful for the financial support from the Stiftung Ravensburger Verlag. We also thank Ludovica Gambaro, Johannes Hauenstein, Josefine Koebe, Jan Marcus, Clara Schäper, and Alexandra Spitz-Oehner for helpful comments and fruitful discussions on the topic. Finally, our thanks goes to the participants at the 7th BIEN 2021 conference in Berlin, the IAAE 2021 annual Conference, the BeNA Summer Workshop 2021, the VfS Annual Conference 2021, the 7th international pairfam User Conference 2021, the 2022 Meeting of SEHO, the European Population Conference 2022, the 14th International German SOEP User Conference 2022 and internal seminars at DIW and BiB.

[2]The term daycare describes all forms of formal childcare provided by professionals outside the family. The term parental childcare describes all childcare provided by the mother or the father of the child. Grandparental care describes the situation in which grandparents take care of their grandchildren, the children in our setting.

variations given country-specific differences in the childcare setting and female labor force participation.[3] In Germany, a country with traditionally low maternal employment and a universal daycare system, every fourth child below the age of eleven is cared for by the grandparents on a regular basis (section 3.3). Although daycare arrangements have expanded over the past decades in Germany as in many other industrialized countries, the relevance of grandparents as informal caregivers has remained relatively stable over the years. The continuously high importance of grandparental care can be attributed to the need to reconcile childcare, (full-time) employment, longer commutes, and non-flexible opening hours of daycare centers, as well as parental and grandparental preferences for this kind of care. Thus, many studies with different foci (for a summary see e.g., Hank and Buber, 2009, section 3.2) have analyzed the relevance of grandparental care in the "care puzzle" of many families, based mainly on US data. However, only a few have a double-generation perspective, looking at both child and parental outcomes – the focus of our paper.

Why is such a perspective interesting? Compared to other caregivers, grandparents might have more time to focus solely on the child. Their greater life experience and emotional closeness might affect children positively in various dimensions. However, if grandparents consider themselves less of a teacher and more of a friend, we might find different effects on socio-emotional skills and school outcomes (e.g., Dunifon et al., 2018). In terms of parental outcomes, grandparental care could provide parents with more time for themselves, leading to improved satisfaction with their leisure time use. In contrast, grandparental care might be accompanied by emotional stress between the grandparents and parents, as inter-familial relationships are more prone to emotional conflicts than those with caregivers outside the family. Compared to other care modes, grandparental care might also be a less stable and continuous care option, for instance, due to sickness or other obligations of the grandparents, which, in turn, could lead to more stress. In general, the intensity of grandparental care – even if it is regular – might not be high enough to substantially affect child and parent outcomes, as this care might be too similar to the alternative care mode. Therefore, it remains an empirical question whether grandparental care pays off. In this paper, we focus on children's health, socio-emotional skills, and school-related outcomes, as well as subjective parental well-being. Both parental and child outcomes are important in the short and medium term, as well as for educational, health, and labor market outcomes at later ages.

While ours is not the only study analyzing such outcomes, it is – to the best of our knowledge – one of the few studies estimating the *causal impact* of grandparental care

[3] Apart from this, more than one third of Europeans see informal care by grandparents or other relatives as the most preferred non-parental care mode (Eurostat, 2012).

on the above-mentioned outcomes. Hereby, we make several contributions. First, we add to the literature on the effects of grandparental care on child outcomes,[4] particularly that on causal effects. The study by Del Boca et al. (2018) focuses exclusively on cognitive outcomes, while Ao et al. (2021) analyze the influence of grandparental care on children's locus of control, based on a sample of three-generation households. While both causal studies focus on one particular child outcome, we focus on a variety of child outcomes.

Secondly, we add to the literature that evaluates the causal impact of regular grandparental care on subjective parental well-being. The study by Chen and Zhang (2018) is one of the few that analyze the causal effect on parental well-being. In comparison, we consider a range of subjective parental well-being outcomes in order to capture potential heterogeneous effects.

Third, we give further evidence of these effects based on data for a country with almost no three-generation households and an increasing share of children in highly subsidized daycare. Thus, we add to the literature, which focuses mainly on the US context or (other) European countries, with different childcare settings and also alternative care modes. Fourth, we account for age-dependent alternative care modes by conducting various subgroup analyses. While for younger children, grandparental care mainly comes on top of sole parental care, for older children, grandparental care is combined with daycare or school visits and parental care (section 3.3). Thus, it gives hints on the role of intergenerational transfers to the next generations on social mobility (e.g., Song and Mare, 2019). We investigate effect heterogeneity and provide some suggestive evidence for the plausible mechanisms behind the effects.

The identification of a causal relationship between grandparental care and child and parental outcomes is difficult because the care decision made by parents and grandparents is endogenous and thus also affects child outcomes. In order to overcome this endogeneity problem, we employ an instrumental variable approach where we use the distance to grandparents as an instrument for grandparental care. The validity of the instrument might be disputable, however, we outline in detail how we make sure that we establish a causal relationship. One major concern could be differences between families living close to the grandparents and families living further away from them. In order to account for this, i.e. to make families that live close and further away com-

[4]Sadruddin et al. (2019) survey 206 studies from more than 50 countries and regions that globally and comprehensively review the impacts of grandparental care on children's outcomes, including physical and mental health, behaviors, cognitive skills, and education. For instance, Fergusson et al. (2008) found that grandparental care was associated with some elevated rates of hyperactivity and peer difficulties at age 4, but these were largely attributable to variations in the types of families using grandparental care. However, they do not claim to find causal relationships.

parable, we combine the IV approach with entropy balancing. Furthermore, we show that neither parents nor grandparents move strategically closer or further away around the time of child birth. We also show that our results are robust to an extensive set of robustness checks concerning the validity of our instrument. Most importantly, we demonstrate that our instrument, the distance between parents/children and grandparents, does not affect our outcome variables per se, but only through care provided by the grandparents by estimating the effects of distance on our outcomes for childless households. Our analysis is based on two representative panel data sets for Germany: *pairfam* and *SOEP*. We use samples of 6,771 and 5,085 families and observe them over a 12-year period (2009-2020) and an 8-year period (2010-2017), respectively. Our analysis relates to children who are usually considered to require some kind of care, namely, children up to the age of ten. Moreover, we focus on regular care in contrast to emergency care by grandparents or other care settings such as during school holidays where parents need support with childcare.

We focus on important outcomes for the next two generations. Cognitive and, to a smaller degree, socio-emotional skills are largely determined early in life (e.g., Cunha and Heckman, 2008). Thus, input provided by carers plays a significant role in child development. Early skills and child health are important preconditions for an effective production of skills in following periods. Moreover, socio-emotional skills promote the formation of school-related outcomes (e.g., Cunha and Heckman, 2007; Rustichini et al., 2017). Child health is equally as important for child development as school outcomes and socio-emotional skills (e.g., Currie, 2020). Parental well-being can be used to measure the utility parents derive from care arrangements and can act as a well-being measure *per se*. Additionally, the well-being of parents affects child development (e.g., Berger and Spiess, 2011; Dahlen, 2016). The interdependence of child and parental outcomes highlights the importance of a double-generation perspective when studying the effects of grandparental care. Parental well-being also influences other important parental outcomes, such as maternal labor supply and fertility (e.g., Sandner, 2019).

Overall, our results provide evidence that, on average, grandparental care does not affect child outcomes; at least, the effects on most of the outcomes we capture are not statistically significant. However, we find that grandparental care negatively affects elementary school children's health, which is mostly driven by children cared for by less healthy grandparents. Concerning parental outcomes, the picture is different, as we find more outcomes to be statistically significantly affected, particularly for maternal well-being. We provide evidence that grandparental care increases maternal and paternal satisfaction with the childcare situation and exhibits positive effects on maternal satisfaction with leisure.

The remainder of this paper is structured as follows: section 3.2 reviews the related literature. In section 3.3 we depict the institutional setting in Germany. Section 3.4 describes the used data set and discusses possible mechanisms of the effects of grandparental care on children and parents. In section 3.5 we present the empirical strategy. Section 3.6 reports the main findings, discusses the robustness of the results and presents the results of our heterogeneity analysis and section 3.7 concludes.

3.2 Contribution to Literature

There is increasing literature on grandparental care in social science, taking different perspectives and approaches, mainly based on US data or European countries other than Germany. Our study contributes to at least three literature strands focusing on the causal relationships[5] of regular grandparental care: studies exploring i) the effects of grandparental care on various outcomes of the grandparents themselves; ii) the effects of various care modes, including grandparental care, on child outcomes; and (iii) the effects of various care modes, again including grandparental care, on parental outcomes.

Causal estimates on the effects of grandparental care on grandparental outcomes, such as health, well-being, and cognitive functioning, are rare and find only limited evidence for a causal association. Danielsbacka et al. (2019) show that positive associations between grandparental care and health and well-being are due only to between-person differences and do not hold in within-person analyses. Arpino and Bordone (2014), however, find positive effects on the verbal fluency of the grandparents but no effects on other cognitive tests. Another paper provides evidence that providing grandparental care leads to a decrease in grandmother's social activities like volunteering (Arpino and Bordone, 2017). A number of studies have shown negative effects of grandparenthood on grandparental labor supply (e.g., Backhaus and Barslund, 2021; Frimmel et al., 2020; Rupert and Zanella, 2018). The effects can be attributed to caring grandmothers who are less attached to the labor market – at least for the cohorts studied so far. We contribute to this literature by focusing on the effects of grandparental care on the care-receiving generations, namely the children and their parents.

The effects of various care modes on child outcomes have been studied extensively in recent years, with a focus on the effects of daycare,[6] while there is hardly any causal

[5] For a recent overview of various studies that mostly analyze these questions as associations, see Hank et al. (2018).

[6] For Germany, see e.g. Bach et al., 2019; Cornelissen et al., 2018; Felfe and Lalive, 2018, who all show positive effects for children from lower socio-economic background in particular, while Kuehnle and Oberfichtner (2020) do not find such effects.

research on the effects of informal care on children. The study by Del Boca et al. (2018) uses UK data to evaluate the effect of grandparental care, instrumented with the distance between the parental and grandparental homes, on cognitive child outcomes at ages 3 to 7, which serve as predictors of school outcomes. Their results suggest that there is no difference in outcomes between children in grandparental care and parental care. However, they find children in grandparental care to be better at naming objects but worse at other skills. Ao et al. (2021) examine the effect of grandparental care on the locus of control of children aged 10 to 15. They use the number of parents' siblings as instrumental variables. With Chinese panel data (CFPS), they find that grandparental care significantly raises children's external locus of control by approximately 1 standard deviation. Thus, children in the care of their grandparents tend to attribute individual success to external factors, such as luck and fate, more than children in parental care. Another study finds that an Austrian parental leave reform crowded out informal care (mostly offered by grandparents) and increased children's cognitive and later labor market outcomes. Danzer et al. (2022) conclude that care provided by mothers is superior to informal care arrangements.[7] We add to this literature by estimating the causal effect of grandparental care on health, socio-emotional, and school outcomes[8] and compare outcomes between children who are in daycare and those who are not, in addition to grandparental care.

The literature on the effects of various care modes, again largely covering daycare or parental care, on parental outcomes is huge and focuses mainly on the effects on maternal employment (for a recent overview, see Müller and Wrohlich, 2020), but also other outcomes such as fertility (e.g., Bauernschuster and Schlotter, 2015; Cools et al., 2015) or maternal well-being.[9] We focus on parental well-being as an outcome that has been studied less extensively.[10] Based on Chinese data, Chen and Zhang (2018) evaluate the causal impact of grandparental retirement (resulting in more potential time for the care of grandchildren) on parental well-being. They find no effect on mothers' subjective health or life satisfaction. We extend this strand of the literature

[7] A study by Milovanska-Farrington (2021) analyzes the relative effects of grandparental supervision compared to parental care time, using Scottish data. Grandparental care time has a positive impact on the observed cognitive skills. However, the causal approach they use applies only to very specific institutional settings.

[8] We use the term "school outcomes" to indicate that the covered measures are not only the results of cognitive skills but non-cognitive skills as well.

[9] While the latter outcome is less investigated, evidence of the effects of daycare in Germany on parental well-being shows mixed but generally positive results (e.g., Kröll and Borck, 2013; Schmitz, 2020; Schober and Schmitt, 2017; Schober and Stahl, 2016).

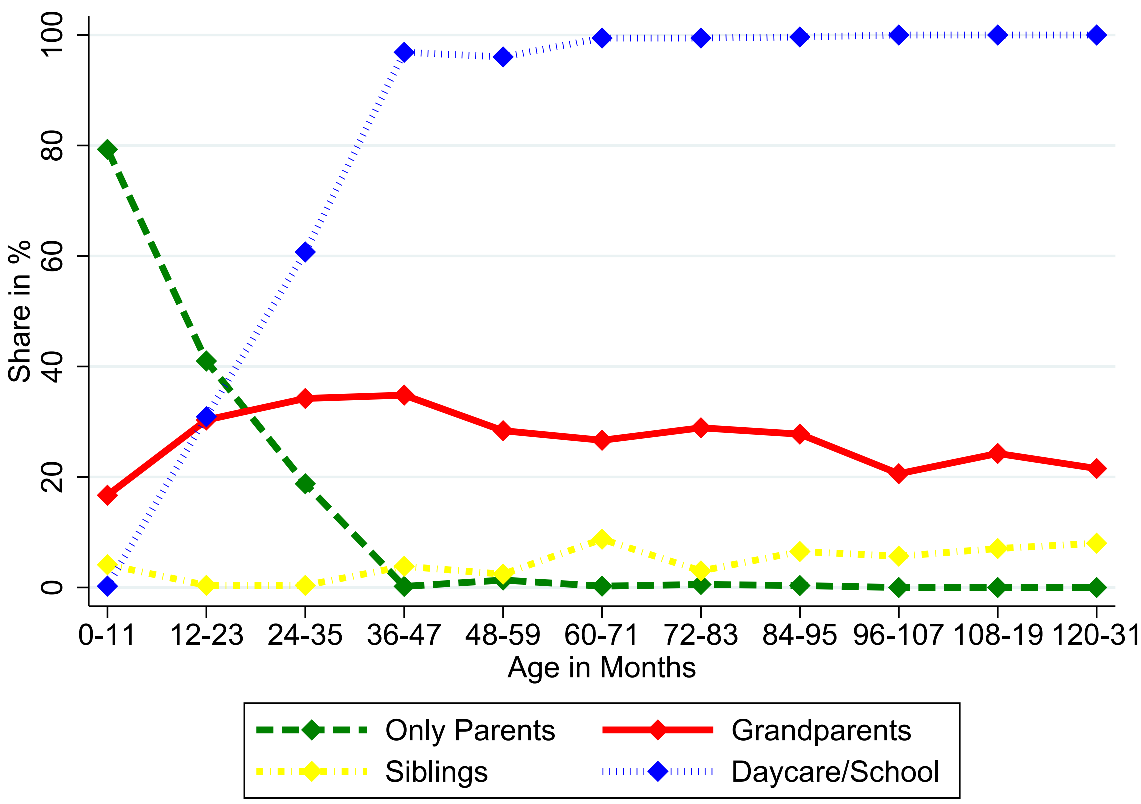
[10] The effects of grandparental care on maternal employment have already been studied quite extensively, showing an increase in maternal employment following grandparental care (e.g., Bratti et al., 2018; Compton and Pollak, 2014; Fenoll, 2020; Kanji, 2018). Also the timing of fertility of parents was shown to be affected by grandparental retirement (Eibich and Siedler, 2020).

by estimating the effect of grandparental care on parental well-being separately for mothers and fathers.

3.3 Institutional setting

In Germany, regular grandparental care has played a significant role for many years (see Figure 3.A.1). Figure 3.1 demonstrates that in 2018/19 across age groups, grandparents cared for about 20 to 30 percent of children below the age of eleven.

Figure 3.1: Actors and institutions involved in care of children younger than 11 in Germany



Notes: The graph shows the share of children cared for by different care actors across age groups. A child is counted as cared for by the grandparents in this graph if the child is cared for by its grandparents in the morning or afternoon or both. The same applies for the other actors.
Source: Pairfam (2018/19), weighted, own calculation.

Over the past decades, maternal employment in Germany has been increasing from 57 percent in 1991 to 72 percent in 2020 (e.g., Destatis, 2022b).[11] This was made possible through a policy that has led to a significant increase in the supply of publicly

[11]In comparison, the average maternal employment rate was 71 percent in 2019 in OECD countries (e.g., OECD, 2020).

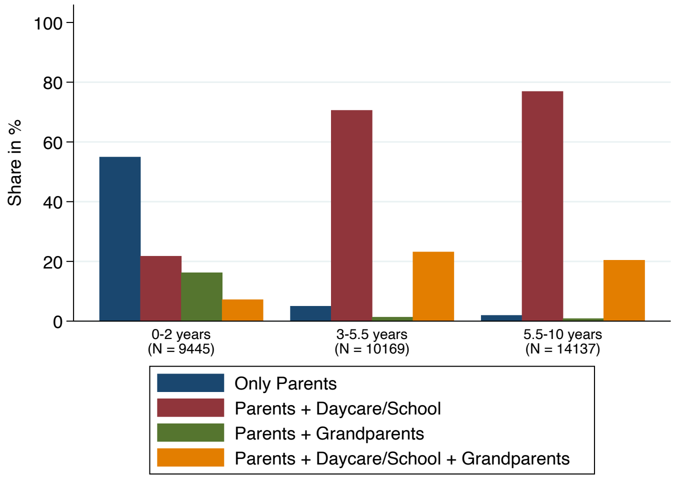
funded daycare since the 1990s (e.g., Müller and Wrohlich, 2020). The proportion of children below the age of three in daycare has seen a substantial increase, from below 5 percent in 1990 to about 29 percent in 2018 (Destatis, 2019b). Still, daycare coverage varies by children's age. Many families with children aged three years and younger do not have a daycare slot, despite the demand (e.g., Jessen et al., 2020). For older children, enrollment has been almost universal (95 percent) since the year 2000 (Destatis, 2019b). However, for this age group there are not enough slots offering full-time care to match parental preferences (Autorengruppe Bildungsberichterstattung, 2020). Daycare fees are relatively low, and some states have even abolished them (e.g., Huebener et al., 2020; Schmitz et al., 2017). The share of for-profit providers is low at about 2 percent (Destatis, 2018). Most daycare centers are operated by non-profit organizations or municipalities. Other forms of regular childcare that have seen a large increase in usage in recent years are all-day schools or after-school care programs. The share of children in all-day schools or related programs increased from 28 percent in 2005/06 to 68 percent in 2018/19. Nevertheless, there is also an excess demand for these slots (Autorengruppe Bildungsberichterstattung, 2020).

Next to formal care arrangements, grandparents play an important role in the "care puzzle." Figure 3.2 shows the share of different combinations of care modes for different child age groups pooled over the period 2009-2020. Panel (a) represents overall care use, taking morning and/or afternoon together, panel (b) shows care use in the morning, and panel (c) care use in the afternoon. The majority of young children (0-2 years) are cared for only by their parents (almost 60 percent). In the morning, the second most frequently used option is a combination of parental and daycare, which applies to about 25 percent of children, followed by a mixture of parental and grandparental care (about 15 percent). In the afternoon, the combination of parental and grandparental care is the second most frequently used option (20 percent), while only about 10 percent of children are cared for by parents and daycare in the afternoon. Thus, we define parental care as the counterfactual (i.e., alternative) care option of grandparental care for this age group.

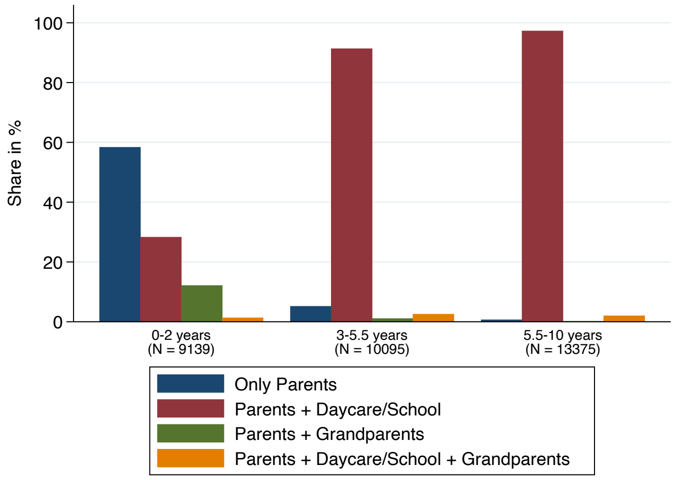
Older children (3-5.5 years and 5.5-10 years, elementary school children) are most frequently cared for by a combination of parents and daycare/school (70-80 percent). Here we observe and expect large differences between morning and afternoon: in the morning, 90-95 percent of children are cared for by either daycare or school, while in the afternoon, only about 30 percent of children are cared for by daycare or school. Here the majority of children are cared for by their parents only (about 50 percent). A substantial number of older children is also cared for by their grandparents in the

Figure 3.2: Care patterns

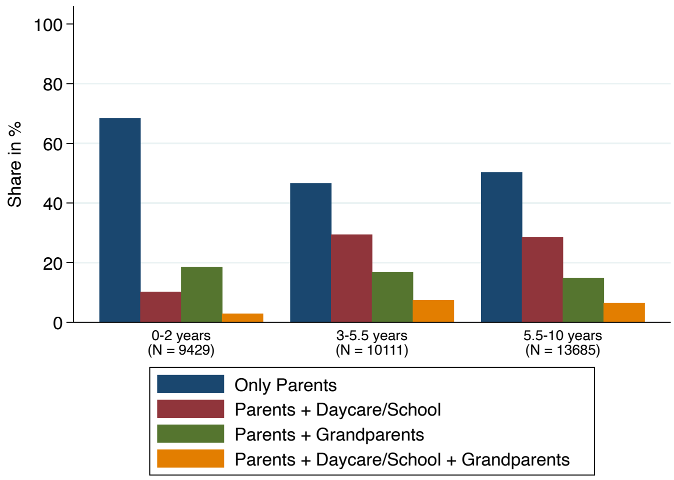
(a) Overall care use by age group



(b) Care use in the morning by age group



(c) Care use in the afternoon by age group



Notes: The figures show the care use by age group. Overall care use takes all actors either caring for the child in the morning or afternoon or both into consideration.
Source: Pairfam (2010-2020), weighted, own calculation.

afternoon: almost 20 percent are cared for by parents and grandparents, and about 10 percent by parents, daycare/school, and grandparents.

In conclusion, the most common *counterfactual* of grandparental care for older children, who are mostly in daycare or school in the morning, is either sole parental care or parental care together with daycare or a school program in the afternoon.

3.4 Data

For the analysis, we use two representative survey datasets that allow us to investigate a large number of different outcomes. The first dataset, which is used to analyze the effects of grandparental care on subjective parental well-being, children's socio-emotional outcomes, and children's health, is the "Panel Analysis of Intimate Relationships and Family Dynamics" (*pairfam*). Participants are surveyed annually (Huinink et al., 2011). We use the *pairfam* data for the information on child health, children's socio-emotional skills, and parental satisfaction measures. To analyze children's school outcomes, we use a second data set. This is the German Socio-Economic Panel *(SOEP)*. The *SOEP* is a representative household and person survey that has been conducted every year since 1984 (Goebel et al., 2019). For more information on the data sets and their comparability, see Appendix 3.A.2.

Grandparental Care Variable. The main explanatory variable in our analysis is the grandparental care variable. In *pairfam*, respondents (parents) are questioned about the regular childcare situation for each child individually. We have information on grandparental care for each child separately for both morning and afternoon, but the data does not allow us to differentiate between grandmothers and grandfathers as caregivers. In the *SOEP*, grandparental care is measured in hours per week. This information is mostly given by the mothers.[12] For the *pairfam*-based analysis, we employ a binary variable that indicates whether a child is regularly cared for by its grandparents in the morning or afternoon or both. To analyze all other parental outcomes, we use a binary variable, which equals one, if at least one child of the parent in question is cared for by the grandparents in the morning or afternoon or both.[13] In the *SOEP*-based analysis, we employ a binary variable, which equals one, if the child is cared for by the grandparents for at least one hour per week. Here we cannot differentiate between morning and afternoon hours. In an additional analysis based on the SOEP data we

[12]However, the hours are not measured for all children, only for particular age groups.

[13]This approximation is valid since in 97 percent of households in our sample, either no or all children are cared for by the grandparents.

also use a continuous variable indicating the number of hours per week a child is cared for by the grandparents.[14]

Child Outcome Variables. We analyze the effects of grandparental care on children's health and developmental skills. To assess the effect on children's health, we consider children's *general health problems*. The *general health* variable is an ordinal variable ranging from 1 (very good health) to 5 (bad health). To estimate the effects of grandparental care on socio-emotional skills, we consider an index variable measuring children's *socio-emotional problems*. This variable in the *pairfam* data is very similar to the internationally widely used SDQ Scale (Strengths and Difficulties Questionnaire, Goodman, 1997). In more detail, we analyze the impact of grandparental care on three indices (conduct problems, hyperactivity, and emotional problems). Summing up the values from these three variables forms the variable *socio-emotional problems*. These questions are asked only for children between 3 and 5 years.

For our analysis of children's school outcomes, we use variables measuring the *Maths* and *German grades* of children between 9 and 10 years. Secondly, the *SOEP* questions mothers about the extent to which the following statements are true: *The child likes going to school* and *The child likes learning*. Both variables are measured on a scale from 1 (strongly agree) to 4 (strongly disagree).[15]

Sample averages of all our outcome measures are shown in Table 3.A.1 in the Appendix. On average, parents rate their children's health as very good: the mean is 1.58, which is close to 1 (very good health). Overall, parents assess the socio-emotional skills of their children as quite high. This is reflected by the relatively low sample mean of the socio-emotional problems variable. Socio-emotional problems are quite evenly distributed across the three components of the socio-emotional problems variable (conduct problems, hyperactivity, and emotional problems). Children in elementary school have, on average, quite good Maths and German grades (2.3) and tend to enjoy going to school and studying. We standardize all child outcomes in our regression analysis in such a way that they have a zero mean and a standard deviation of one.

Parental Outcome Variables. We use several variables on subjective parental satisfaction. We consider six variables, which are all ordinal variables on an 11-point Likert scale ranging from 0 (very dissatisfied) to 10 (very satisfied). The first variable captures the general satisfaction with life. In addition, *pairfam* contains several variables on domain-specific satisfaction. First, we consider the satisfaction with school, education, or career. Secondly, respondents are asked about their satisfaction with

[14]The variance of this variable is quite small, 67% of children are in grandparental care for less than 3 hours.

[15]All four variables are surveyed from 2012 onwards.

leisure activities, hobbies, and interests. Thirdly, they are asked to rate their satisfaction with the relationship with their partner. And from 2013 onwards, they are also asked about their satisfaction with their work-life balance. Furthermore, parents are asked about their satisfaction with the childcare situation for each of their children. Thus, we can analyze the effect for each child separately.

Sample means pooled across age groups are shown in Table 3.A.1 in the Appendix. For most outcomes, mothers and fathers depict similar levels of satisfaction. Interestingly, the levels of satisfaction are also similar across the various domains. Overall, individuals in our sample show quite high levels of satisfaction ranging between 5.9 and 8.5.

Measurement of the instrument. We employ the distance to grandparents as an instrument for grandparental care, which is further explained in section 3.5. Both data sets include information on the geographical distance between the adults in the household and all four grandparents (if they are still alive) in several waves. In *pairfam*, we measure this by comparing families that live 30 minutes or less from the grandparents to those living further away, while in the *SOEP*, the instrument compares living in the same city to living in another city. In more detail, in *pairfam*, the distance to the grandparents is part of the "anchor questionnaire" as well as the partner questionnaire and is asked in all waves with the following question: “How much time do you need to get to your mother’s dwelling? (on a normal day, using normal means of transportation)”. In cases in which the anchor’s or partner’s parents do not live in one household, they are asked the same question about the distance to the father’s dwelling. The distance is measured as a categorical variable with six categories.[16] Based on this, we construct a binary variable which equals unity if at least one grandparent lives closer than 30 minutes and equals zero otherwise. We employ this binary variable because the relationship between the distance and the amount of grandparental care provided is unlikely to be linear. For example, the difference between living 10 or 30 minutes away should have a larger impact than the difference between 3 hours and 3 hours and 20 minutes. We use 30 minutes as the cut-off, as this is a reasonable distance that still allows commuting within one day when giving care to a grandchild.[17] The distribution of the ordinal distance variable used to construct our instrument and the grandparental care variable in *pairfam* can be seen in Figure 3.A.2 in the Appendix. This figure shows the share of children that are in grandparental care by the minimum

[16]The six categories are: “we live in one house”, “less than 10 minutes”, “10 minutes to less than 30 minutes”, “30 minutes to less than 1 hour”, “1 hour to less than 3 hours” and “3 hours or more”.

[17]In robustness checks, we test whether our results are sensitive to two different definitions of the instrument (using an ordinal instrument and using one hour as the cutoff). See Appendix section 3.A.4.

distance of the child to the grandparents. It can be seen that most children live close to at least one grandparent. In the whole sample, about 69 percent of households live less than 30 minutes away from at least one grandparent.[18] Additionally, it can be seen that the share of households that use grandparental care increases non-linearly with decreasing distance.[19]

In the *SOEP*, the distance to the grandparents is surveyed in the parents' individual questionnaires using the following question: "Which and how many of the following relatives do you have? Please also state where they live." The distance is measured as a categorical variable with the seven categories, which is surveyed every 5 years.[20] In our analysis, we use the distance obtained in 2011 and 2016. In order to use a larger sample for our analysis, we impute the distance in the year before and after it was surveyed. This means that our *SOEP* analysis is based on the years 2010-2012 as well as 2015-2017. Just as for *pairfam*, we define a binary variable of the distance which equals unity if at least one grandparent lives in the same town as the household (but more than 15 minutes away by foot) and 0 otherwise. 52 percent of households in the *SOEP* sample live in the same town as at least one of the grandparents.

Control Variables. To account for other observable factors that might confound the effect of grandparental care on child outcomes and family well-being, our models include extensive sets of control variables on the (grand-)parental, child, and household level. Generally, we include socio-economic characteristics of the parents, such as education, age, income, labor force status, gender, federal state of residence, and migration background. Additionally, we include detailed information about the situation of the household (e.g., number of children in the household and age of the youngest child). An overview of the set of control variables for each outcome variable is given in Table 3.A.2 in the Appendix. In robustness checks, we vary the set of included control variables (e.g., excluding potentially endogenous variables such as maternal labor forces status and income) and show that our results are robust to these changes.

Samples. We conduct analyses on the child and parent level. To evaluate the effects on child outcomes and parental satisfaction with the childcare situation, each child constitutes one observation. The analysis sample for all other parental outcomes

[18] This percentage is weighted and based on the child data set of *pairfam*. In the parental level data set, 70 percent of households live closer than 30 minutes away from at least one grandparent.

[19] It appears that of those households in our sample that live further than three hours away from all grandparents, slightly more than 5 percent still report using grandparental care on a regular basis. As this seems unlikely, we exclude those households in a robustness check, which does not change our results. The results are available from the authors upon request.

[20] The seven categories are: "here in this same household", "in the same house, but in another household", "in the same neighborhood", "in the same town, but more than 15 minutes away by foot", "in another town, but within a one hour drive", "further away, but in Germany", and "abroad."

is restricted to all individuals who have at least one child in the appropriate age group. These analyses are conducted at the parent level. Furthermore, we restrict our analysis to families in which at least one parent was born in Germany. If they were both born outside Germany, it is highly likely that all four grandparents do not live in Germany and are therefore not available for regular childcare (e.g., Gambaro et al., 2018). We observe both *pairfam* samples from 2009 to 2020[21] and the *SOEP* sample from 2010 to 2012 and 2015 to 2017. Our final sample to analyze socio-emotional and child health outcomes includes 44,339 observations, which corresponds to 11,714 children. The sample to analyze school outcomes includes 34,904 observations, which corresponds to 9,047 children. The analysis sample for parental outcomes, using *pairfam*, includes 16,056 observations for fathers (corresponding to 4,043 fathers) and 19,844 observations for mothers (corresponding to 4,788 mothers).

3.5 Empirical strategy

In order to identify the causal effect of grandparental care on the various outcomes under study, we apply an instrumental variable (IV) strategy. In a simple OLS setting, the regression model would look like this:

$$y_{it} = \beta_1 + \beta_2 GPC_{it} + X'_{it}\beta_3 + \mu_{it} \tag{3.1}$$

where y_{it} are the different child and parent outcome variables. The variable of interest, grandparental care (GPC_{it}), is a binary variable, and X'_{it} is our vector of control variables, as described in section 3.4. However, employing the OLS model in Equation 3.1 does not necessarily produce estimates that can be interpreted causally. The identification of a causal effect of grandparental care on child and parental outcomes faces potential endogeneity threats. The choice for grandparental care is endogenous as it is made by parents and grandparents and might be influenced by unobserved characteristics that also influence the outcome variables, causing an omitted variable bias. One example of such an unobserved variable is a grandparent's preferences for taking care of their grandchild. These likely influence the amount of support grandparents offer and might also directly affect our outcomes. Another threat could be reverse causality; for example, parental well-being might influence how much support from the grandparents they need and thus demand. Similarly, children's health or socio-emotional problems are likely to affect the decision to ask grandparents for help. For example, parents with children who suffer from bad health might fear that taking care of these children would

[21] For 2020, we include only households that were surveyed before March 15 and thus before the beginning of the COVID pandemic in Germany.

be too much of a burden for grandparents or they really need the grandparents as no other non-parental care mode is feasible.

Thus, estimating Equation 3.1 might lead to a biased and inconsistent estimator of grandparental care and would not reflect a causal effect. There are reasons to expect both upward biased and downward biased OLS estimators. For example, if only healthy and socio-emotionally stable children are in grandparental care, we expect the OLS estimator to be upward biased. Alternatively, if we expect that parents with low subjective well-being are more likely to ask grandparents for childcare assistance because they are more in need of help, the OLS estimator would be downward biased. We cannot account for the endogeneity issues by including all confounding factors as control variables as some of them are not observed in the data at hand or might be unknown.

To overcome the endogeneity problem, we use an instrumental variable, applying a two-stage least-squares (2SLS) approach. We can predict the variation in grandparental care using an instrument that determines the endogenous regressor (GPC_{it}) but only affects the dependent variables (y_{it}) through its effect on this independent variable (grandparental care). For that purpose, we use the distance to the grandparents as an instrument. This instrument was also used by Del Boca et al. (2018) and Compton and Pollak (2014).

Validity of the instrument. In order for the distance to grandparents to qualify as a valid instrument, it needs to fulfill a number of conditions. Particularly important are the relevance and the exogeneity assumptions of the instrument. Relevance means that the instrument needs to be sufficiently correlated with the endogenous regressor grandparental care. Arguably, the distance to the grandparents satisfies the relevance condition as a smaller distance facilitates grandparental care. The correlation between our instrument and grandparental care can be seen in Figure 3.A.2 in Appendix 3.A.2. This figure shows the share of children who are in grandparental care by the minimum distance of the child to the grandparents. Additionally, it can be seen that the share of households that use grandparental care increases non-linearly with decreasing distance. The correlation between the instrument and the endogenous regressor is also tested in the first stage regression where the endogenous variable is regressed on the instruments and the exogenous covariates (Table 3.1). The robust first stage F-statistics displayed in the main regression tables in section 3.6 are all at least 55 but far exceed this value in most regressions. This supports our argument.[22]

[22] We tested three further potential instruments using a pension reform in Germany, the parents' birth order, and the gender of the oldest sibling of both parents. All three instruments proved to be weak instruments (small first stage F-statistic).

The more critical assumption is the exogeneity assumption of the instrument, which requires that the instrument is not correlated with the error term and thus influences the outcome variable only through the endogenous regressor. It seems plausible that distance affects child outcomes only through grandparental care. It can be argued, however, that living close to the grandparents affects parental well-being not only through grandparental care but also through the relationship to the grandparents and the amount of time parents and grandparents can spend together. To ensure that distance only affects parental outcomes through the grandparental care provided, we control for the emotional closeness between parents and grandparents in a robustness check. Furthermore, it can be argued that childcare demand increases the probability of families living closer to the grandparents (e.g., Chen and Zhang, 2018). To further test the exogeneity of the distance to the grandparents, we investigate whether distance between parents and grandparents decreases around birth, which would indicate that either parents moved closer to the grandparents or grandparents moved closer to the parents. The reason for a systematic moving behavior could be the facilitation of grandparental childcare, which would make distance an endogenous variable. Investigations of the moving behavior in the year before and after the birth of the first-born or any child show no systematic movement towards the grandparents (see Table 3.A.12). We further restrict the sample to households that did not move during the observation period, thus excluding any households that might have moved closer to the grandparents in order to facilitate childcare. However, the results do not change (see Tables 3.A.13 and 3.A.14).

To test whether households that live close to the grandparents and households that live further away differ in their characteristics, we regress the distance dummy on our control variables (as described in section 3.4). Education, labor force status, migration background, number of children and participation in early education services seem to be predictors of the distance (Table 3.A.15). These characteristics being associated with choice of the location of residence are in line with findings of Siedentop et al. (2014). In order to account for the differences between families living close and further away, we combine our IV estimation with entropy balancing (Hainmueller, 2012), a matching strategy that balances controls more effectively than propensity score methods. We first conduct this matching step and then run our regular IV estimations. The main idea of entropy balancing is to assign a weight to observations in the "control group" (families living further away than 30 minutes) causing the "control group's" distributions of the selected covariates to match those of the "treatment group" (families living closer than 30 minutes) on the mean. Consequently, our set of covariates have the same means in both groups. These weights are then applied to our IV estimations. We discuss the results in the robustness section. For more details on the measurement and the validity of the instrument, see the data and the robustness sections.

Two-Stage Least Squares. In the first stage of our 2SLS approach, we regress the grandparental care variable that we assume to be endogenous on our instrument and the exogenous control variables:

$$GPC_{it} = \gamma_1 + \gamma_2 D_{it} + X'_{it}\gamma_4 + \varepsilon_{it} \tag{3.2}$$

where D_{it} equals one if the household lives less than 30 minutes away from at least one grandparent and 0 otherwise[23] and X'_{it} is the same vector of control variables as in Equation 3.1. The dependent variable GPC_{it} is the binary grandparental care variable from Equation 3.1. The first stage regression is estimated using OLS. Since the dependent variable is binary, this corresponds to a linear probability model (LPM, see Appendix 3.A.4). In a further robustness check, we also conduct a probit estimation (called a "garden variety") as suggested by Angrist and Pischke (2008) (see Appendix 3.A.4). In the second stage, the fitted values of the linear probability model from the first stage $\widehat{GPC}_{it}$ are included as the main explanatory variable:

$$y_{it} = \beta_1 + \beta_2 \widehat{GPC}_{it} + X'_{it}\beta_3 + \mu_{it} \tag{3.3}$$

In this regression, y_{it} are the different child and parental outcome variables described in section 3.4. X'_{it} is again our vector of control variables that is the same as in the first stage regression. β_2 is our coefficient of interest and reflects the 2SLS estimator. Per definition it estimates the local average treatment effect (LATE)[24] and thus depicts the effect of grandparental care on our outcomes. In our case there are no always-takers as living far away prevents regular grandparental care. Therefore, our 2SLS estimator reflects the average treatment effect on the treated (ATT).[25]

[23]For the analyses based on the *SOEP*, this is defined as 1 for households living in the same city as the grandparents and 0 otherwise.

[24]It measures the effect on the compliers, i.e., those families whose utilization of grandparental care is induced by a small distance to the grandparents.

[25]The robust standard errors μ_{it} are clustered at the household level for all regressions using child outcomes and the parental satisfaction with the childcare situation because the observations of different children in one household might be correlated with each other and, as a result, the i.i.d. assumption would not hold. Clustering at the household level allows individuals to be correlated within households and across time. Robust standard errors are used for all other parental outcomes.

3.6 Empirical Results

We start the discussion on the effects of grandparental care with a discussion on the first-stage effects. For all outcomes, the effects of distance on grandparental care are highly significant and of similar magnitude (Table 3.1). Living at a maximum of half an hour from at least one grandparent leads to an increase in the probability of grandparental care by about 23 percentage points (depending on the outcome). This suggests that our instrument is very relevant, i.e., there is a high correlation between instrument (distance) and the endogenous variable (grandparental care).[26]

Table 3.1: First stage results

Health & Socio-emotional skills:	Health	Socio-emot. problems	Conduct	Hyperactivity	Emotional	
Distance	0.254***	0.239***	0.239***	0.239***	0.239***	
	(0.018)	(0.029)	(0.029)	(0.029)	(0.029)	
R-squared	0.130	0.161	0.161	0.161	0.161	
Observations	11069	2171	2172	2173	2172	
School outcomes:	Math grade	German grade	Child likes going to school	Child likes studying		
Distance	0.320***	0.320***	0.289***	0.289***		
	(0.038)	(0.038)	(0.031)	(0.031)		
Observations	1475	1476	2278	2261		
R-Squared	0.207	0.207	0.187	0.188		
Parental Satisfaction:	General	Educ./ career	Leisure	Relationship	Work-life balance	Child care
Distance: Maternal Sat.	0.233***	0.236***	0.238***	0.236***	0.242***	0.285***
	(0.024)	(0.013)	(0.013)	(0.013)	(0.013)	(0.020)
Observations	5838	6182	6061	6182	5742	2514
R-Squared	0.147	0.149	0.149	0.149	0.152	0.200
Distance: Paternal Sat.	0.245***	0.239***	0.239***	0.239***	0.239***	0.269***
	(0.033)	(0.016)	(0.016)	(0.016)	(0.016)	(0.023)
	4011	4495	4490	4494	4491	2510
R-Squared	0.183	0.159	0.159	0.159	0.160	0.188
	4,481	4,476	4,480	4,477	2,504	4,011

Notes: Standard errors in parentheses. Conditional on no missings in the outcome and control variables (see Table 3.A.2).
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Next, we discuss the effects of grandparental care on child outcomes.

Child Outcomes. The upper panel of Table 3.2 displays the effects on children's health and socio-emotional behavior. General health problems are analyzed for four different age groups. Remember that the counterfactual to grandparental care varies by age group. While for the majority of children younger than three years of age, the

[26]The first-stage results are not sensitive regarding the choice of control variables as shown in Tables 3.A.16 and 3.A.17 in the Appendix.

counterfactual is sole parental care, this is different for older children. For them, the counterfactual is either half-daycare or school and sole parental care in the afternoon or full-time daycare and school combined with parental care.

As high values in the general health variable correspond to bad health, the coefficient for health problems (all children, row one in the upper panel) suggests that grandparental care has a negative effect on the health of children below the age of 11 (column 2). The effect is statistically significant on the 5 percent level: grandparental care increases children's health problems by 0.46 standard deviations. This corresponds to a 29 percent increase compared to the sample mean. The effect seems to be mostly driven by children of elementary school age as the coefficient of this subsample estimation is of similar magnitude and significance to the coefficient for all children. For children in the other age groups, the coefficient is not significant.

Table 3.2 also allows the comparison of the OLS and IV estimates. We note that the OLS estimate (column 1) underestimates the effect of grandparental care on health for all age groups. While not significant and very small in magnitude, the OLS estimates indicate smaller negative effects (or even positive effects) on health for children in grandparental care than the IV estimator. This finding supports our hypothesis that parents with children with bad health tend not to ask grandparents for help.

The effects of grandparental care on children's socio-emotional problems are displayed in rows five to eight in the upper panel of Table 3.2. The direction of the IV estimates suggests that grandparental care increases socio-emotional problems of children aged 3–5 (the only age group for which we have this measure). However, all effects are statistically not significant. A comparison with the OLS estimates shows that the pure correlations are positive and statistically significant, meaning that grandparental care is associated with a decrease in the socio-emotional problems of children. This hints that there might be a bias in the way that parents of more socio-emotionally stable children use grandparental care more often.

The lower panel of Table 3.2 depicts the effects of grandparental care on children's school outcomes. Although the IV estimates suggest a deterioration in the Math grade, an improvement in the German grade, an increase in the willingness to go to school, and a decrease in the willingness to study following grandparental care, all effects are statistically not significant. This is also true for the OLS estimates, which all suggest positive associations of grandparental care and school-related skills. We can conclude that grandparental care has no impact on the children's school-related skills, at least the ones we capture.

Table 3.2: Effects of Grandparental Care on Child Outcomes

	Grandparental Care OLS	Grandparental Care IV	F-Statistic	Sample Mean	Obs.
Health					
Health problems: 0-10 y.	0.017	0.464*	198.819	1.574	11069
	(0.039)	(0.183)			
Health problems: 0-2 y.	-0.014	0.484	68.817	1.546	1828
	(0.068)	(0.348)			
Health problems: 3-5.5 y.	-0.039	0.254	118.187	1.579	3006
	(0.054)	(0.194)			
Health problems: 5.5-10 y.	0.057	0.438*	155.568	1.573	5132
	(0.051)	(0.194)			
Socio-emotional behavior					
Socio-emotional problems: 3-5 y.	-0.142**	0.365	70.350	2.943	2171
	(0.049)	(0.275)			
Conduct problems: 3-5 y.	-0.030	0.217	70.490	1.064	2172
	(0.053)	(0.303)			
Hyperactivity: 3-5 y.	-0.161**	0.275	70.690	1.002	2173
	(0.056)	(0.251)			
Emotional problems: 3-5 y.	-0.132*	0.331	70.690	0.878	2172
	(0.053)	(0.279)			
School outcomes					
Math grade: 9-10 y.	-0.138	0.0459	77.930	2.264	1476
	(0.092)	(0.188)			
German grade: 9-10 y.	-0.136	-0.124	78.127	2.300	1477
	(0.093)	(0.220)			
Child likes going to school: 9-10 y.	0.078	-0.014	98.428	1.556	2262
	(0.065)	(0.208)			
Child likes studying: 9-10 y.	0.105	0.183	98.371	1.924	2245
	(0.071)	(0.199)			

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

If we further differentiate by (all-)daycare status, to account for differences in the counterfactual care modes (parental care vs. parental care plus school/daycare), Table 3.3 shows a slightly different picture. If children three years and older are in daycare or school full time and, in addition, cared for by grandparents, they have more health and socio-emotional problems, in particular, conduct problems. This might be related to a greater instability of caregivers in the afternoon, which might be too stressful for some children as they have to deal with various caregivers in various care settings during one afternoon (e.g., Bratsch-Hines et al., 2015). Additionally, these children like studying less than those who are not in additional grandparental care. Comparably, children who are in half-daycare show more health problems once they are in grandparental care in the afternoon, but no difference in socio-emotional problems, which underlines our hypothesis that too many care modes might increase behavioral problems.

Parental Outcomes. The effects of grandparental care on parental satisfaction are shown in Table 3.4. The results for mothers are summarized in the upper panel and for fathers in the lower panel. The IV estimates (column 2) of grandparental care on the maternal satisfaction outcomes displayed are all positive, suggesting that grandparental care increases maternal satisfaction. More precisely, the table depicts statistically significant effects for maternal satisfaction with both the childcare situation and leisure time. The effects correspond to an increase of 11 percent for satisfaction with the childcare situation and 14 percent for satisfaction with leisure compared to the mean (column 4). A comparison of the IV and OLS estimates shows that for all maternal satisfaction outcomes, the OLS estimator underestimates the effects of grandparental care. One explanation for this could be that parents with generally low well-being require help and thus make more use of grandparental care.

Finally, we analyze how grandparental care affects paternal satisfaction, measured with the same variables as maternal satisfaction. As for mothers, grandparental care increases fathers' satisfaction with the childcare situation statistically significantly, while the effect is substantially larger in magnitude. The increase corresponds to approximately 21 percent compared to the mean. Additionally, childcare provided by the grandparents decreases fathers' satisfaction with their career and education by 7 percent in comparison to the mean. However, this effect is only significant at the 10 percent significance level. The remaining well-being measures are not significantly affected by grandparental care.

We further estimate effects for different child age groups and different counterfactual care modes to get a more precise picture of the driving forces of the effects. The estimates for satisfaction with the childcare situation are significant at the 10 percent

Table 3.3: Child outcomes by daycare status

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
(a) 0-2 years: Child in daycare				
Health problems: 0-2 y.	0.190 (0.522)	29.347	1.651	587
(b) 0-2 years: Child not in daycare				
Health problems: 0-2 y.	0.430 (0.402)	44.533	1.503	1241
(c) 3-10 years: Child in daycare/school full-time				
Health problems: 3-10 y.	0.550^{+} (0.307)	71.919	1.583	2762
Socio-emotional problems: 3-5 y.	1.061^{+} (0.551)	32.981	2.966	971
Conduct problems: 3-5 y.	1.170^{*} (0.586)	33.183	1.146	972
Hyperactivity: 3-5 y.	0.805 (0.496)	33.183	0.989	972
Emotional problems: 3-5 y.	0.357 (0.437)	32.981	0.832	971
Math grade: 9-10 y.	0.463 (0.351)	31.177	2.294	405
German grade: 9-10 y.	0.301 (0.347)	31.177	2.286	405
Child likes going to school: 9-10 y.	-0.334 (0.380)	27.569	1.502	631
Child likes studying: 9-10 y.	0.911^{*} (0.413)	25.280	1.889	627
(d) 3-10 years: Child in daycare/school part-time				
Health problems: 3-10 y.	0.346^{+} (0.202)	145.495	1.572	5295
Socio-emotional problems: 3-5 y.	0.205 (0.292)	35.359	2.928	1200
Conduct problems: 3-5 y.	0.085 (0.342)	35.359	1.012	1200
Hyperactivity: 3-5 y.	0.162 (0.281)	35.569	1.010	1201
Emotional problems: 3-5 y.	0.215 (0.344)	35.569	0.907	1201
Math grade: 9-10 y.	-0.133 (0.201)	51.274	2.263	1040
German grade: 9-10 y.	-0.222 (0.230)	51.425	2.314	1041
Child likes going to school: 9-10 y.	-0.035 (0.247)	68.942	1.587	1591
Child likes studying: 9-10 y.	-0.125 (0.223)	70.355	1.953	1578

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.4: Effects of Grandparental Care on Parental Satisfaction

Outcomes	Grandparental Care OLS	IV	F-Statistic	Sample Mean	Obs.
Mother's Satisfaction with:					
Child care situation	0.118	0.922^{*}	98.205	8.481	5838
	(0.094)	(0.463)			
Life	0.015	0.041	328.912	7.759	6182
	(0.047)	(0.212)			
Education, Career	0.088	0.396	324.348	7.171	6061
	(0.067)	(0.293)			
Leisure, Hobbies	0.035	0.892^{**}	328.769	6.325	6182
	(0.070)	(0.308)			
Relationship to Partner	0.116	0.214	327.011	7.561	5742
	(0.071)	(0.313)			
Work-life Balance	-0.242^{*}	0.130	208.277	6.429	2514
	(0.108)	(0.383)			
Father's Satisfaction with:					
Child care situation	0.334^{**}	1.761^{***}	55.698	8.496	4011
	(0.109)	(0.527)			
Life	0.025	0.198	220.800	7.802	4495
	(0.048)	(0.203)			
Education, Career	0.052	-0.511^{+}	220.158	7.494	4490
	(0.060)	(0.275)			
Leisure, Hobbies	-0.101	-0.066	221.138	6.451	4494
	(0.071)	(0.316)			
Relationship to Partner	-0.006	-0.252	220.281	7.681	4491
	(0.078)	(0.354)			
Work-life Balance	-0.099	-0.374	141.937	5.903	2510
	(0.107)	(0.426)			

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

significance level for mothers with children aged 3-5.5 years (Table 3.A.4, panel (b)). The estimates suggest an increase that corresponds to 15 percent compared to the mean. The effect on satisfaction with leisure is largely due to mothers with children of elementary school age (5.5 to 10 years, panel (c)) and very young children (0 to 2 years, panel (a)). The first effect is highly statistically significant and corresponds to a 24 percent increase compared to the sample mean.

For fathers, we find more statistically significant effects by child age (Table 3.A.5). The estimates for satisfaction with the childcare situation are at least significant on the 10% significance level across all age groups (panel (a) - (c)) and especially large in magnitude for children below the age of 3. Fathers with very young children are also more satisfied with their life once grandparents support. We find a negative effect of grandparental care on the satisfaction with work-life balance and education and career for fathers with children 3-5.5. years of age (10 and 21 percent decreases, respectively). However, these effects are not robust (see Appendix 3.A.4).

If we further differentiate by (all-)daycare status, to account for differences in the counterfactual care modes, Tables 3.5 and 3.6 show the following: The increase in satisfaction with leisure mainly stems from mothers whose infants are not in daycare or whose older children are not in full-time daycare/school. Once older children are in full-time care/school and additional grandparental care, mothers are even less satisfied with their life and their relationship to their partner – maybe because this also produces more stress for them as well as for the children (see above). This is different if their children are only in half-daycare. This leads to an increase in satisfaction with both the care situation and leisure. For fathers, the results differ: the increase in life satisfaction and satisfaction with the childcare situation of infants comes from fathers of infants who are in daycare.

Heterogeneity. We did further subsample analyses by parental education, gender of the child, grandparental health and grandparental age and discuss how these could reflect potential mechanisms through which grandparental care has an impact on children and parents. We enrich our analysis of heterogeneous treatment effects by estimating causal forests (Wager and Athey, 2018). Thereby, we get a better understanding of the treatment effects at different points of the (grandparental/child) age distribution.

First, as it is known from the literature that there are differences in child outcomes by child gender, we estimate different models for boys and girls. The negative health effects can be mostly attributed to boys as the coefficient is larger in magnitude and statistically more significant (Table 3.A.6). In terms of school outcomes, there is a

Table 3.5: Mother's Satisfaction by daycare status

Mother's Satisfaction with:	IV: GPC	F-Statistic	Sample Mean	Obs.
(a) 0-2 years: Child in daycare				
Child care situation	1.666 (1.285)	9.482	8.484	244
Life	-0.539 (0.561)	31.314	7.753	665
Education, Career	0.142 (0.755)	31.762	7.142	654
Leisure, Hobbies	-0.198 (0.765)	31.314	5.897	665
Relationship to Partner	-0.178 (0.862)	29.472	7.562	634
Work-life Balance	-1.482^{+} (0.860)	22.470	6.256	295
(b) 0-2 years: Child not in daycare				
Child care situation	1.295 (1.065)	19.917	8.691	536
Life	0.579 (0.445)	63.583	8.002	1453
Education, Career	1.062 (0.711)	59.292	7.059	1381
Leisure, Hobbies	1.853* (0.788)	63.583	6.158	1453
Relationship to Partner	0.610 (0.639)	63.588	7.794	1397
Work-life Balance	-0.292 (1.244)	14.080	6.408	233
(c) 3-10 years: Child in daycare/school full-time				
Child care situation	-0.00548 (0.866)	33.256	8.354	1448
Life	-0.784* (0.396)	95.172	7.468	1761
Education, Career	0.520 (0.507)	96.641	7.128	1744
Leisure, Hobbies	0.369 (0.517)	95.172	6.050	1761
Relationship to Partner	-1.651** (0.586)	93.804	7.301	1608
Work-life Balance	0.441 (0.703)	61.780	6.054	909
(d) 3-10 years: Child in daycare/school part-time				
Child care situation	1.139* (0.524)	86.117	8.512	2929
Life	0.459 (0.310)	153.166	7.778	3109
Education, Career	0.195 (0.430)	152.277	7.204	3049
Leisure, Hobbies	1.107* (0.443)	153.043	6.391	3109
Relationship to Partner	0.476 (0.443)	161.240	7.572	2909
Work-life Balance	-0.339 (0.671)	60.229	6.636	1187

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.6: Father's Satisfaction by daycare status

Father's Satisfaction with:	IV: GPC	F-Statistic	Sample Mean	Obs.
(a) 0-2 years: Child in daycare				
Child care situation	3.166^{+} (1.865)	4.693	8.341	192
Life	1.787^{*} (0.808)	19.288	7.823	612
Education, Career	-0.551 (0.750)	19.341	7.580	610
Leisure, Hobbies	0.783 (1.017)	19.288	6.359	612
Relationship to Partner	-0.279 (0.861)	19.081	7.679	611
Work-life Balance	-1.338 (2.138)	5.618	5.719	357
(b) 0-2 years: Child not in daycare				
Child care situation	2.963 (2.885)	5.131	8.849	379
Life	0.247 (0.361)	62.492	7.969	1244
Education, Career	-0.314 (0.495)	62.514	7.496	1244
Leisure, Hobbies	0.159 (0.603)	62.462	6.274	1244
Relationship to Partner	-0.502 (0.619)	62.469	7.856	1244
Work-life Balance	1.405 (0.899)	31.197	5.928	670
(c) 3-10 years: Child in daycare/school full-time				
Child care situation	2.256^{*} (0.972)	15.278	8.197	1045
Life	0.253 (0.486)	36.147	7.742	1273
Education, Career	-0.164 (0.612)	35.512	7.456	1272
Leisure, Hobbies	-0.823 (0.706)	36.147	6.394	1273
Relationship to Partner	-1.080 (0.836)	36.432	7.555	1269
Work-life Balance	-2.978^{*} (1.178)	23.392	5.904	745
(d) 3-10 years: Child in daycare/school part-time				
Child care situation	1.242^{*} (0.593)	51.841	8.587	1937
Life	0.246 (0.285)	132.069	7.713	2182
Education, Career	-0.086 (0.381)	131.863	7.468	2180
Leisure, Hobbies	0.054 (0.432)	132.398	6.460	2181
Relationship to Partner	0.375 (0.493)	131.979	7.670	2179
Work-life Balance	-0.520 (0.547)	111.363	5.863	1147

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

marginally significant reduction in “child likes studying” for boys who are in grandparental care.

Secondly, we evaluate the effect of grandparental care on children’s health by grandparental health. In line with our prior expectations, grandparental care has a negative effect on children’s health when their own health is equal to or below median health (3.A.7). Grandparents with worse health are likely to be physically restricted and therefore might conduct fewer activities that include movement with their grandchildren (e.g., fewer walks and outdoor activities). This could contribute to a worse health status of children.

Thirdly, differentiating by parental education, the estimations reveal that the negative health effects can be mostly attributed to children of parents who hold at least one university degree. For all other child outcomes, there are no notable differences between children with parents who hold a university degree and children with parents who do not hold a university degree (Table 3.A.8). The positive effect on mothers’ satisfaction with childcare is more pronounced for mothers who hold a university degree than for mothers who do not (Table 3.A.9). This could be explained by the fact that more highly educated mothers usually work more hours and therefore have more of a problem reconciling childcare and work duties without the help of grandparents. In contrast, the positive effect on satisfaction with leisure is about twice as large for mothers without a university degree. One reason could be that grandparents support more highly educated mothers with reconciling childcare and work while they give less educated mothers the chance to reconcile work, childcare, and leisure time. For fathers, the picture looks different. While the positive effect on satisfaction with childcare can be mostly attributed to fathers who do not hold a university degree, the negative effect on satisfaction with education and career is more significant for more educated fathers (Table 3.A.10). One explanation could be that grandparental care is a less reliable care option than, for example, daycare, and thus more highly educated fathers feel hampered in their career development.

Lastly, we separate the sample at the median grandparental age (about 64 years). Table 3.A.11 shows that health problems caused by grandparental care are most prominent among children that are cared for by grandparents below median age. This could be explained by grandparental employment, i.e., grandparents below median age are likely still active on the labor market, while older grandparents are likely retired. Retired grandparents might have more time for outdoor activities or to cook healthy meals.[27]

[27] As these effects are also visible for more healthy grandparents, the age effects can likely be attributed to older but healthy grandparents.

To better understand the heterogeneous treatment effects we estimate *causal forests*[28] and graphically present the predicted treatment effects by child and grandparental age.[29] Figure 3.A.3, panel (a) confirms the results we find for the effects on child health, namely that the effects are similar across all age groups and particularly pronounced for older children. Panel (b) shows a decreasing trend of the predicted treatment effect on child health by grandparental age. This also coincides with our finding from Table 3.A.11, that the negative effect on childrens' health is driven by younger grandparents. Finally, plots (c) and (d) also show a decreasing predicted treatment effect with increasing age of the child. This is in line with the results found in Tables 3.A.4 and 3.A.5.

Robustness. To further corroborate our findings and test the exogeneity of the instrument used, we conduct several robustness checks. Some robustness checks concerning the validity of the instrument (e.g., analysis for childless households or using the distance to the individual's parents-in-law) we conduct only for parental outcomes (see above). It can be argued that the distance to the grandparents likely affects child outcomes only through the time spent with the grandparents, i.e., grandparental care. For parents, this relationship is less straightforward, but we prove through several robustness checks that we are able to isolate the effect of grandparental care on parental satisfaction.

For example, we apply entropy balancing as described in section 3.5 prior to running the IV estimations for both child and parental outcomes. By doing so, we equalize differences in observables between families that live close and further away from their grandparents. The results are shown in Tables 3.A.18 and 3.A.19. The effects remain very similar, we still depict highly significant effects on children's health, parental satisfaction with the childcare situation and maternal satisfaction with leisure. However, the negative effect on paternal satisfaction with education and career is no longer statistically significant.

Next, we use only the distance to the individual's parents-in-law (instead of the distance to any grandparent) as an instrument when estimating the effects of grandparental care on parental outcomes. The idea behind this is that the relationship beyond childcare is usually closer to one's own parents than to one's parents-in-law (e.g., Del Boca et al., 2018). Thus, in case the distance to the own parents has some effect on parental satisfaction through some factor other than childcare that we cannot

[28]We use the R-package *grf* and estimate an instrumental forest.

[29]Note, graphically representing heterogeneous treatment effects is particularly interesting for continuous variables. Thus, we focus in this analysis on continuous variables where we found significant differences in our heterogeneity analyses, namely child age for child health and satisfaction with childcare for mothers and fathers and grandparental age for child health.

control for, this should be ruled out when using the distance to the parents-in-law. The results show that our instrument proved to be a strong instrument, measured by the first stage F-statistic. Generally, the results are similar to our main results (see Table 3.A.20). As in the robustness check using entropy balancing, the effect on paternal satisfaction with career is no longer statistically significant.

Additionally, we estimate the same regressions as in the main analysis for childless households. With this analysis, we provide further evidence that our specification isolates the effects of grandparental care on parental well-being, i.e., we control for all other channels through which distance affects parental well-being. In more detail, as "grandparents"[30] in childless households do not provide childcare, the estimates capture the effect of distance on well-being other than childcare. Thus, if our main analysis isolates the effect of grandparental childcare on well-being, estimating the same equation for childless households should not identify any effects of distance to the “grandparents” on parents' well-being. Table 3.A.21 shows that the point estimates are very small in magnitude and that there are no statistically significant effects of distance on well-being for both childless women and childless men.[31]

Furthermore, we include further control variables, namely, emotional closeness of parents and grandparents, frequency of contact between parents and grandparents, grandparental health, and pre-birth satisfaction values of parents and exclude potential bad controls (income and labor force status) to prove the robustness of our results. The results are shown in Tables 3.A.30 and 3.A.31. The results are very robust to the change in the set of control variables. For a more detailed description see Appendix 3.A.4.

Further robustness checks (e.g. a placebo analysis, correcting for multiple hypothesis testing or adding/excluding control variables) are provided in Appendix 3.A.4. Overall, the results on parental satisfaction with childcare and maternal satisfaction with leisure are most robust, while those on child health and paternal satisfaction with career should be interpreted with caution, at least for the overall sample.

[30] These are the parents or parents-in-law of childless adults, they correspond to our grandparent generation.

[31] Because individuals in childless households are, on average, younger than parents in households with children in *pairfam* (the mean age of childless individuals is 29.95, and that of our baseline sample is 36.36), we exclude the youngest quartile of the sample in additional regressions in order to make the childless sample more comparable to our main sample. In these analyses, we still do not find any effects of the distance on well-being.

3.7 Conclusion

With our analysis, we contribute to the literature on the intergenerational effects of regular grandparental care on outcomes of parents and children. Our results are of particular interest as grandparental care continues to play an important role in the "care puzzle". This development will probably not change as the overlap of lifetimes of the child, parent, and grandparent generations is increasing with increasing longevity. We extend the literature on grandparental care by estimating the causal effects on health, socio-emotional and school-related outcomes of children and parental well-being. To overcome endogeneity between grandparental care and our outcomes, we employ an instrumental variable approach instrumenting grandparental care with the distance to the grandparents, which we combine with entropy balancing. We show various robustness checks supporting the validity of our instrument.

Using two representative panel data sets, our results for the overall sample provide evidence for mainly null and a few negative effects on children and mainly positive effects on different aspects of parental satisfaction. However, the results differ widely according to child age. Regarding the average null effects on socio-emotional and school outcomes of children, one might argue that grandparental care is neither beneficial nor costly for the grandchildren generation. Regarding child health and older children, it is partly costly, although we focus only on short-term effects. This is different for the generation of parents. Here, grandparenting is beneficial at least for maternal well-being. Thus, it might also be beneficial for the child's development in the longer run, as maternal well-being has been found to positively impact child outcomes. This might be an indirect effect on the grandchildren generation and thus might affect overall social mobility.

Specifically, we find evidence for a negative effect of grandparental care on the health of elementary school children (20 percent).[32] The health effect is particularly pronounced for the sample cared for by less healthy grandparents. Results of studies on the health effects of other care modes, such as daycare, are mixed. Cornelissen et al. (2018) find positive health effects of daycare that are similar in magnitude to our effects. Namely, they depict a 25 percent decrease in "compensatory sports needed" at school entry. Baker et al. (2008) find negative health effects of a major daycare expansion in Canada, which amount to 9 percent compared to the mean. Given that this is the first causal evidence on the effect of grandparental care on child health, there is no comparison with other estimates possible.

[32] However, as our estimate turns out to be less significant in some of our robustness checks, we interpret this effect with some caution.

We do not find overall effects of grandparental care on socio-emotional skills of children. However, once the sample is restricted to older children in full-time daycare or school, we find that additional grandparental care increases socio-emotional problems. Baker et al. (2008) also find that daycare increases children's anxiety-related emotional disorder score by 12 percent. Datta Gupta and Simonsen (2010) find enrollment into family homecare in Denmark increases the SDQ index by 28 percent, which corresponds to an increase in adverse behavior, while Peter et al. (2016) find a decrease in the SDQ when children in the UK visit daycare early.

Our results on school outcomes show hardly any significant and causal relationship, with the exception that 9-10-year-olds who are cared for by their grandparents in the afternoon, in addition to full-time schooling, like studying less than those without additional grandparental care. The insignificance of the effects on school grades is in line with the findings of Del Boca et al. (2018): while they find some effects on school-related outcomes of children below school age, they find no effects for children once they have entered elementary school.

The positive effects of grandparental care on parents' satisfaction with childcare, as well as mothers' satisfaction with leisure, are very robust to different specifications, sample restrictions, and instruments. The negative effects found for fathers' satisfaction with their education and career turn out to be less robust and thus should be interpreted with caution. Comparing our effects with the effects of daycare attendance on maternal life satisfaction as, for instance, depicted by Schmitz (2020), shows that our effects (11-14 percent) are larger in magnitude. Schmitz (2020) finds an 8 percent increase in comparison to the mean.

Overall, our results show that not only parental care and daycare affect child and family outcomes, but that regular childcare provided by other informal caregivers, such as grandparents, also has causal impacts on children and parents and thus the family as a whole. However, we also have only suggestive evidence on the mechanisms behind these effects. To investigate them, data that cover the activities grandparents do with their grandchildren would be needed (e.g. Sadruddin et al., 2019). Moreover, as with other care modes, more information on the quality of the care time would be needed (Milovanska-Farrington, 2021). And finally, longer-term effects should be investigated to analyze whether the positive effects on maternal satisfaction increase child outcomes and other maternal outcomes and thus grandparental care has additional indirect effects.

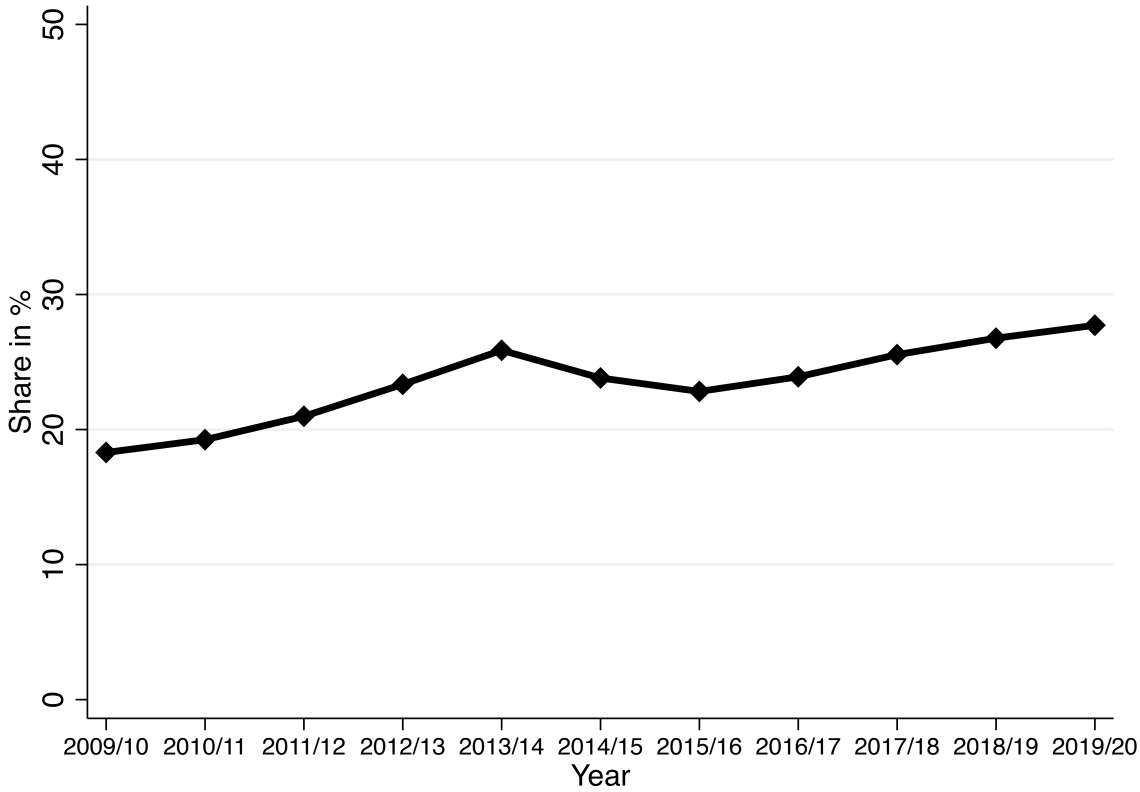
From a policy perspective, it should be clear that a focus not only on daycare but also on informal care is needed. For instance, there could be discussions on national insurance credits for grandparents who take care of dependent children, contributing to their retirement income, as implemented in the UK. Another measure to support grandparental care might be the introduction of grandparental leave and benefits[33] as in Portugal (Milovanska-Farrington, 2021), or “grandparenting allowances” (e.g., Wheelock and Jones, 2002). Nevertheless, our results also suggest that the combination of too many care modes might have negative effects on children and parents. Politicians might address this by policies that are in favor of longer daycare hours or other measures to reduce the “child penalty” employed parents might have if the opening hours of daycare centers do not support their working schedules (e.g., Jessen, 2022).

[33]Since 2008, in particular circumstances, grandparents of children, for instance with mothers younger than 18, can apply for parental leave in Germany.

3.A Appendix

3.A.1 Graph on grandparental care

Figure 3.A.1: Development of grandparental care (2009-2020)



Notes: The graph shows the development of grandparental care for children below the age of 6. A child is counted as cared for by the grandparents if the child is cared for by its grandparents in the morning or afternoon or both.
Source: Pairfam (2010-2020), weighted, own calculation.

3.A.2 Further information on the data

Pairfam. *Pairfam* respondents are equally distributed among the birth cohorts 1991–1993, 1981–1983, and 1971–1973 and the first wave of the sample consisted of 12,400 respondents (Huinink et al., 2011). These individuals are called "anchor persons." Approximately one half of the anchors are male, and the other half are female. In addition, if anchors and anchors' partners agreed, partners were surveyed from the first wave onwards. The response rate for partners lies at about 52 percent.[34] *Pairfam* is a multi-actor survey. In addition to anchors and partners, children (aged 8 to 15 years) and parents of anchors are surveyed separately. Furthermore, anchors and partners are questioned about their children (biological, adopted, foster, and stepchildren of anchors living in one household) and parents in their own questionnaires in detail (Huinink et al., 2011). This detailed information on three generations makes *pairfam* particularly suitable for our analysis. Since the child survey only includes children above the age of 7 and the parent survey suffers from a low response rate, we focus on the information obtained from the anchor and partner questionnaires in our analysis. However, *pairfam* covers no school-related questions. For these outcomes, we use the *SOEP*.

SOEP. The *SOEP* currently surveys about 15,000 households and 30,000 individuals (Goebel et al., 2019). It includes information about all individuals living in one household. In addition to individual questionnaires filled out by all adults in the household, there is a household questionnaire that includes questions on all children living in the household and age-specific child questionnaires which are mostly answered by the mother of the child. In contrast to *pairfam*, grandparents themselves are surveyed only if they live in the same household as the family or if our "parent" used to be a child in a *SOEP* household and has now formed their own household. Thus, the sample for which detailed information on the grandparents is available is a small and very specific sample, which is why we do not use it.

Comparability of Pairfam and SOEP. Table 3.A.3 includes summary statistics of selected control variables for both *pairfam* (based on the sample on child level) and *SOEP*. Columns 1 *(Pairfam)* and 2 *(SOEP)* show mean and standard deviation for selected control variables across all observations. Comparing the two data sets suggests differences in socio-economic characteristics. Moreover, the share of children in grandparental care in the *SOEP* is almost twice as high as in *pairfam*.[35] This might be due to the differences in the phrasing of the question and the way grandparental care is measured (see section 3.4). The *pairfam* sample is, on average, more highly educated, as the share of households in which at least one partner holds a university

[34] Analyses show that anchors whose partners participate and anchors whose partners do not participate do not differ systematically in most of their socio-economic characteristics. Thus, the partner sample can be considered as good as random.

[35] In the *pairfam* wave 12, parents of school children are only questioned about care arrangements in the afternoon. Thus, we defined school children in wave 12 to be cared for by grandparents only if they are cared for by them in the afternoon. This means that there is a very small share of children that are cared for by the grandparents in the morning before school that are counted as not in grandparental care if they are not also in grandparental care in the afternoon. Figure 3 shows that this is only a very small share of school children.

degree is about 12 percentage points higher than in the *SOEP* (37 percent vs. 49 percent).[36] In terms of migration background, household income, age of children and mothers, gender of the children, and number of children in the household, the samples are quite comparable. The differences in socio-economic characteristics emphasize the importance of including our extensive set of control variables as mentioned above. Moreover, we discuss various subsample analyses to show the effect heterogeneity by child, parent, and grandparent characteristics.

[36] Generally, *pairfam* includes a slightly more highly educated sample than the German population (Wetzel et al., 2021).

Table 3.A.1: Sample means of outcome variables

Health & Socio-emotional probl.:	Health problems	Socio-emot. problems	Conduct	Hyperactivity	Emotional	
Children	1.580 (0.694)	3.280 (2.247)	1.141 (1.017)	1.133 (1.034)	1.006 (0.917)	
Observations	25,138	5,078	5,088	5,085	5,085	
School outcomes:	Math grade	German grade	Child likes going to school	Child likes studying		
Children	2.259 (0.829)	2.301 (0.828)	1.563 (0.706)	1.937 (0.816)		
Observations	1,479	1,480	2,283	2,266		
Satisfaction:	General	Educ./ career	Leisure	Relationship	Work-life balance	Child care
Mother	7.759 (1.580)	7.169 (2.142)	6.325 (2.136)	7.561 (2.124)	6.431 (2.210)	8.481 (1.878)
Observations	6,174	6,053	6,174	5,736	2,512	5,838
Father	7.802 (1.369)	7.495 (1.710)	6.449 (1.908)	7.679 (2.086)	5.898 (2.096)	8.496 (1.606)
Observations	4,481	4,476	4,480	4,477	2,504	4,011

Notes: Standard deviations in parentheses. Conditional on no missings in the control variables. Conduct problems, hyperactivity and emotional problems are each constructed by summing two variables that range between 0 (does not apply) and 2 (fully applies). Therefore, conduct problems, hyperactivity and emotional problems range between 0 and 4 and socio-emotional problems between 0 and 12. Note, the questions for socio-emotional problems and health are phrased negatively, meaning that high values correspond to negative characteristics.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.2: Control variables

			Effects on				
			Children's			Parents'	
Variable	Definition	Type	(a)	(b)	(c)	(d)	(e)
Parental Variables							
Post-secondary education	Highest degree in household, 1-3	Ord	✓	✓	✓		
	Individual education, 3 levels	Ord				✓	✓
Mother's labor force status	Parental level, 1-3	Ord	✓	✓	✓	✓	✓
Father's labor force status	Parental level, 1-3	Ord	✓	✓		✓	✓
Age	Mother's age	Cont	✓	✓	✓		
	Individual age	Cont				✓	✓
Religion	One parent religious	Bin	✓	✓			
	Individual religion, 1-7	Cat				✓	✓
Migration background	One parent has direct background	Bin	✓	✓	✓		
	Individual has direct background	Bin				✓	✓
Partner information	Partner answered questionnaire	Bin	✓	✓			
Parental goals	Importance nutrition and exercise, 1-10	Ord	✓				
Health	At least one parent is sick	Bin	✓	✓	✓		
	Individual health, 1-5	Ord				✓	✓

Continued on the next page

Table 3.A.2 continued

Variable	Definition	Type	Effects on Children's (a)	(b)	(c)	Parents' (d)	(e)
Obesity	At least one parent is obese	Bin	✓	✓			
	Individual is obese	Bin				✓	✓
Pregnancy	Parent is pregnant	Bin	✓	✓		✓	✓
Cohabitation	Parents live together	Bin	✓	✓	✓	✓	✓
Widowhood	One parent is widowed	Bin	✓	✓	✓	✓	
	Individual is widowed	Bin					✓
Only child	At least one parent is only child	Bin	✓	✓			
	Individual is only child	Bin				✓	✓
Satisfaction childcare	On the child level, 1-10	Ord		✓			
Child Variables							
Sex	Child's sex	Bin	✓	✓	✓	✓	
	Children in HH: male, female, mixed	Cat					✓
Child age	In months	Cont	✓	✓	✓	✓	
	Age of youngest child in months	Cont					✓
Number children in HH	Total	Cont	✓	✓	✓	✓	
	Nr. children 0-2 years	Cont					✓
	Nr. children 3-5 year	Cont					✓
	Nr. children 6-10 year	Cont					✓
	Nr. other children	Cont					✓
Birth order	Age in comparison to sibling's age	Ord	✓	✓	✓	✓	
Daycare use	Child (0-5 years) in daycare	Bin	✓	✓		✓	
	Number of children (0-5 years) in daycare	Cont					✓
Health	Child health, 1-5	Ord		✓		✓	
	Mean health children, 1-5	Ord					✓
Temperament	Child 0-6 years, 1-20	Ord		✓			
Grandparent Variables							
School education	Anchor's mother, 1-3	Ord	✓	✓		✓	✓
	Anchor's father, 1-3	Ord	✓	✓		✓	✓
	Mother's mother, 1-5	Ord			✓		
	Mother's father, 1-5	Ord			✓		
	Fathers's mother, 1-5	Ord			✓		
	Fathers's father, 1-5	Ord			✓		
Age	Mean of all available grandparents	Cont	✓	✓	✓	✓	✓

Continued on the next page

Table 3.A.2 continued

Variable	Definition	Type	Effects on Children's (a)	(b)	(c)	Parents' (d)	(e)
Household (HH) Variables							
Household income	logarithmic, in 1000 €	Cont	✓	✓	✓	✓	✓
Year	number according to wave number	Cont	✓	✓	✓	✓	✓
Federal state	1-16	Cat	✓	✓	✓	✓	✓
Community size	1-7	Ord	✓	✓	✓	✓	✓

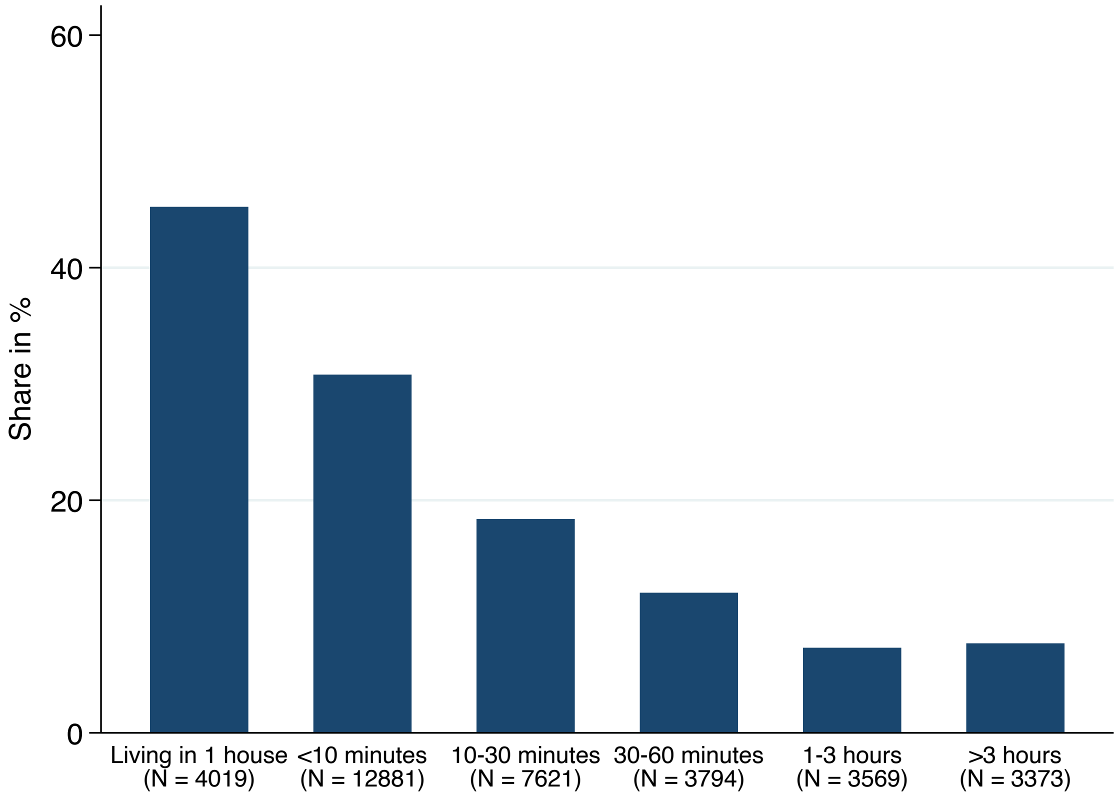
Notes: This table shows which variables are used to estimate the effect of grandparental care on: (a) Child's health problems (b) child's socio-emotional behavior (c) child's school outcomes (d) Parental satisfaction with childcare (e) Other parental satisfaction outcomes. Types: Bin (binary), Cat (categorical), Cont (continuous), Ord (Ordinal).
Source: Pairfam, 2009-2019 (columns a, b, d, e). SOEP, 2010-2012 and 2015-2017 (column c).

Table 3.A.3: Summary Statistics

	Pairfam: Mean (SD) Year: 2009-2020	SOEP: Mean (SD) Year: 2005-2017
Grandparent care	23.323 %	45.860 %
Grandparent live 30 min or closer/ in the same city or closer	68.852 %	51.856 %
Mother's labour force status (in percent)		
Mother not working	36.332 %	43.159 %
Mother working part-time	42.669 %	43.472 %
Mother working full-time	18.879 %	13.369 %
Household's highest parental school degree (in percent)		
No/ lower secondary degree	5.923 %	6.429 %
Upper secondary/vocational degree	45.509 %	55.933 %
University degree	48.569 %	37.639 %
One parent has migration background	11.899 %	12.304 %
Household net income (in Euro)	3416.561 (2430.786)	3298.097 (1850.606)
Age mother (in years)	34.024 (7.898)	36.286 (6.007)
Sex child: male	50.880 %	52.380 %
Number of children in household	2.043 (0.989)	1.989 (0.915)
Age child (in years)	4.904 (3.101)	4.885 (3.173)
Cohabitation with partner	91.068 %	81.816 %
Observations	29,169	12,690

Notes: Means and standard deviations of selected control variables conditional on non-missing sample.
Source: Pairfam 2010-2020, SOEP (2010-2017) weighted, own calculations.

Figure 3.A.2: Grandparental care by distance



Notes: The figures show the share of children cared for by grandparents by the distance between the child's household and the closest living grandparent. A child is counted as cared for by the grandparents in this graph if the child is cared for by its grandparents in the morning or afternoon or both.
Source: Pairfam (2010-2020), weighted, own calculation.

3.A.3 Further subsample analyses

Table 3.A.4: Mother's Satisfaction by child age

Mother's Satisfaction with:	IV: GPC	F-Statistic	Sample Mean	Obs.
(a) Age: 0-2 years				
Child care situation	1.256 (0.923)	26.015	8.632	780
Life	0.305 (0.361)	100.176	7.929	2118
Education, Career	0.869 (0.546)	95.600	7.084	2035
Leisure, Hobbies	1.107^{+} (0.603)	100.176	6.082	2118
Relationship to Partner	0.313 (0.507)	99.883	7.727	2031
Work-life Balance	-1.024 (0.629)	53.664	6.326	528
(b) Age: 3-5.5 years				
Child care situation	1.279^{+} (0.719)	45.992	8.485	1543
Life	-0.053 (0.340)	115.951	7.750	2341
Education, Career	0.304 (0.465)	114.736	7.235	2289
Leisure, Hobbies	0.496 (0.479)	116.202	6.175	2340
Relationship to Partner	-0.689 (0.512)	114.713	7.479	2211
Work-life Balance	1.058 (0.811)	41.232	6.252	898
(c) Age: 5.5-10 years				
Child care situation	0.637 (0.497)	86.297	8.454	2864
Life	0.246 (0.298)	166.064	7.675	3270
Education, Career	0.504 (0.410)	166.042	7.154	3221
Leisure, Hobbies	1.526*** (0.441)	165.696	6.358	3271
Relationship to Partner	0.155 (0.439)	172.102	7.489	3015
Work-life Balance	-0.133 (0.551)	115.585	6.519	1505

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.5: Father's Satisfaction by child age

Father's Satisfaction with:	IV: GPC	F-Statistic	Sample Mean	Obs.
	(a) Age: 0-2 years			
Child care situation	3.093^{+} (1.860)	10.128	8.698	571
Life	0.581^{*} (0.292)	100.432	7.928	1856
Education, Career	-0.512 (0.384)	101.059	7.520	1854
Leisure, Hobbies	0.398 (0.479)	100.787	6.298	1856
Relationship to Partner	-0.232 (0.459)	100.046	7.806	1855
Work-life Balance	0.662 (0.741)	48.788	5.865	1027
	(b) Age: 3-5.5 years			
Child care situation	1.518^{*} (0.691)	43.082	8.410	1082
Life	0.0121 (0.295)	91.457	7.766	1835
Education, Career	-0.770^{*} (0.379)	90.816	7.511	1833
Leisure, Hobbies	-0.355 (0.467)	91.096	6.295	1833
Relationship to Partner	0.301 (0.456)	91.012	7.607	1831
Work-life Balance	-1.209^{+} (0.725)	41.097	5.827	960
	(c) Age: 5.5-10 years			
Child care situation	1.629^{**} (0.577)	45.635	8.496	1926
Life	0.297 (0.318)	95.372	7.704	2200
Education, Career	0.125 (0.440)	94.822	7.454	2199
Leisure, Hobbies	-0.223 (0.455)	95.513	6.525	2200
Relationship to Partner	-0.107 (0.540)	95.074	7.653	2196
Work-life Balance	-0.830 (0.625)	65.009	5.888	1241

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.6: Child outcomes by gender

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
(a) Boys				
Health problems: 0-10 y.	0.586* (0.250)	116.167	1.615	5616
Socio-emotional problems: 3-5 y.	0.921 (0.587)	17.492	3.013	1081
Conduct problems: 3-5 y.	0.490 (0.589)	17.578	1.111	1082
Hyperactivity: 3-5 y.	0.862 (0.543)	17.752	1.004	1083
Emotional problems: 3-5 y.	0.696 (0.563)	17.667	0.898	1082
Math grade: 9-10 y.	0.043 (0.306)	29.709	2.153	758
German grade: 9-10 y.	-0.109 (0.343)	30.068	2.453	759
Child likes going to school: 9-10 y.	0.229 (0.355)	40.557	1.670	1151
Child likes studying: 9-10 y.	0.567+ (0.318)	39.906	2.046	1142
(b) Girls				
Health problems: 0-10 y.	0.389+ (0.208)	135.468	1.532	5453
Socio-emotional problems: 3-5 y.	0.221 (0.293)	66.442	2.874	1090
Conduct problems: 3-5 y.	0.122 (0.332)	66.442	1.017	1090
Hyperactivity: 3-5 y.	0.191 (0.262)	66.442	0.999	1090
Emotional problems: 3-5 y.	0.190 (0.285)	66.442	0.859	1090
Math grade: 9-10 y.	0.013 (0.222)	63.421	2.380	718
German grade: 9-10 y.	-0.252 (0.230)	63.421	2.140	718
Child likes going to school: 9-10 y.	-0.119 (0.231)	75.893	1.440	1111
Child likes studying: 9-10 y.	-0.038 (0.214)	75.798	1.796	1103

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.7: Child outcomes by grandparents' health

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
(a) Health better than median				
Health problems: 0-2 y.	0.680 (0.789)	4.833	1.527	182
Health problems: 3-5.5 y.	0.328 (0.433)	16.852	1.486	245
Health problems: 5.5-10 y.	0.092 (0.576)	5.422	1.422	323
Health problems: 0-10 y.	0.220 (0.333)	30.433	1.482	806
(b) Health worse than/equal to median				
Health problems: 0-2 y.	0.241 (0.530)	16.852	1.454	264
Health problems: 3-5.5 y.	0.822 (0.574)	14.842	1.551	383
Health problems: 5.5-10 y.	0.530^{+} (0.310)	36.473	1.543	528
Health problems: 0-10 y.	0.577^{*} (0.280)	56.733	1.526	1285

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.8: Child outcomes by education

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
(a) University Degree				
Health problems: 0-10 y.	0.608** (0.234)	117.091	1.522	6525
Socio-emotional problems: 3-5 y.	0.374 (0.320)	38.439	2.847	1359
Conduct problems: 3-5 y.	0.278 (0.380)	38.602	1.144	1360
Hyperactivity: 3-5 y.	0.324 (0.302)	38.602	0.922	1360
Emotional problems: 3-5 y.	0.236 (0.326)	38.439	0.782	1359
Math grade: 9-10 y.	-0.175 (0.192)	64.493	1.929	471
German grade: 9-10 y.	-0.144 (0.203)	64.493	1.948	471
Child likes going to school: 9-10 y.	0.130 (0.211)	85.061	1.499	699
Child likes studying: 9-10 y.	0.0944 (0.208)	83.823	1.800	693
(b) No University Degree				
Health problems: 0-10 y.	0.0554 (0.311)	79.387	1.638	4544
Socio-emotional problems: 3-5 y.	0.485 (0.690)	19.262	3.082	812
Conduct problems: 3-5 y.	0.715 (0.660)	19.262	0.948	812
Hyperactivity: 3-5 y.	0.201 (0.656)	19.379	1.117	813
Emotional problems: 3-5 y.	0.149 (0.676)	19.379	1.017	813
Math grade: 9-10 y.	0.201 (0.273)	41.974	2.420	1005
German grade: 9-10 y.	-0.0348 (0.316)	42.166	2.464	1006
Child likes going to school: 9-10 y.	-0.0277 (0.305)	42.379	1.581	1563
Child likes studying: 9-10 y.	0.416 (0.299)	42.539	1.977	1552

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.9: Mother's Satisfaction by education

Mother's Satisfaction with:	IV: GPC	F-Statistic	Sample Mean	Obs.
	(a) University degree			
Child care situation	1.862^{*} (0.727)	40.040	8.476	3299
Life	0.260 (0.225)	201.317	7.956	2366
Education, Career	0.573^{+} (0.302)	200.893	7.623	2313
Leisure, Hobbies	0.735^{*} (0.356)	201.317	6.381	2366
Relationship to Partner	0.524^{+} (0.318)	196.732	7.847	2274
Work-life Balance	0.458 (0.452)	145.222	6.292	1092
	(b) No University degree			
Child care situation	0.416 (0.713)	61.238	8.488	2539
Life	-0.148 (0.366)	121.545	7.658	3816
Education, Career	0.303 (0.536)	118.080	6.939	3748
Leisure, Hobbies	1.139^{*} (0.527)	121.370	6.295	3816
Relationship to Partner	0.0639 (0.584)	115.828	7.408	3468
Work-life Balance	-0.419 (0.677)	59.182	6.514	1422

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.10: Father's Satisfaction by education

Father's Satisfaction with:	IV: GPC	F-Statistic	Sample Mean	Obs.
	(a) University degree			
Child care situation	1.276* (0.650)	42.755	8.487	2510
Life	0.323+ (0.192)	207.703	7.919	2231
Education, Career	-0.534* (0.250)	206.808	7.756	2229
Leisure, Hobbies	-0.271 (0.306)	207.495	6.352	2230
Relationship to Partner	0.508+ (0.304)	206.655	7.743	2229
Work-life Balance	-0.328 (0.437)	122.833	5.909	1314
	(b) No University degree			
Child care situation	2.256+ (1.348)	15.712	8.508	1501
Life	-0.176 (0.460)	46.277	7.699	2264
Education, Career	-0.732 (0.669)	46.222	7.263	2261
Leisure, Hobbies	0.623 (0.722)	46.797	6.538	2264
Relationship to Partner	-1.017 (0.795)	46.420	7.626	2262
Work-life Balance	-0.321 (0.945)	27.383	5.896	1196

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

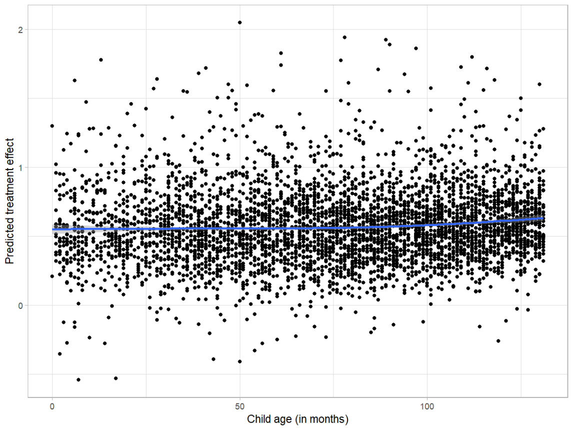
Table 3.A.11: Child outcomes by grandparents age

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
(a) Age above median				
Health problems: 0-10 y.	0.436* (0.222)	126.849	1.548	5997
Socio-emotional problems: 3-5 y.	0.473 (0.443)	21.573	2.811	879
Conduct problems: 3-5 y.	0.570 (0.484)	21.573	1.088	879
Hyperactivity: 3-5 y.	-0.134 (0.386)	21.573	0.928	879
Emotional problems: 3-5 y.	0.667 (0.452)	21.573	0.796	879
(b) Age below than/equal to median				
Health problems: 0-10 y.	0.725** (0.271)	89.584	1.606	5072
Socio-emotional problems: 3-5 y.	0.356 (0.463)	27.955	3.064	837
Conduct problems: 3-5 y.	0.224 (0.479)	28.046	1.035	838
Hyperactivity: 3-5 y.	0.626 (0.446)	28.188	1.071	839
Emotional problems: 3-5 y.	-0.107 (0.461)	28.097	0.959	838

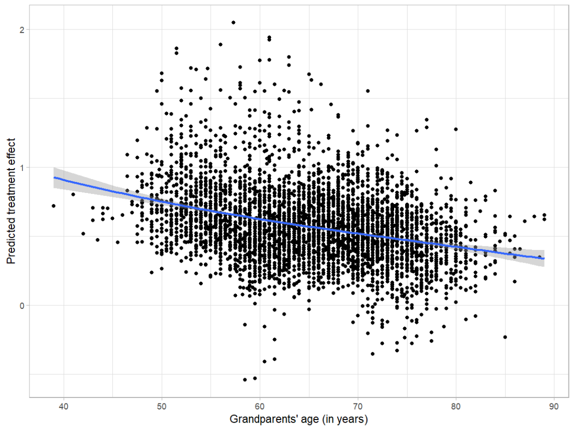
Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Figure 3.A.3: Causal forest predictions

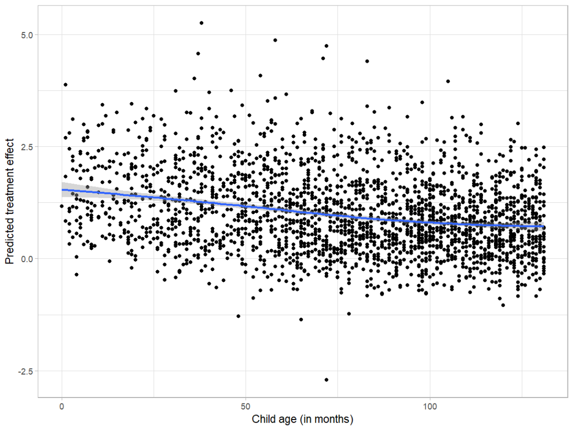
(a) Child health by child age



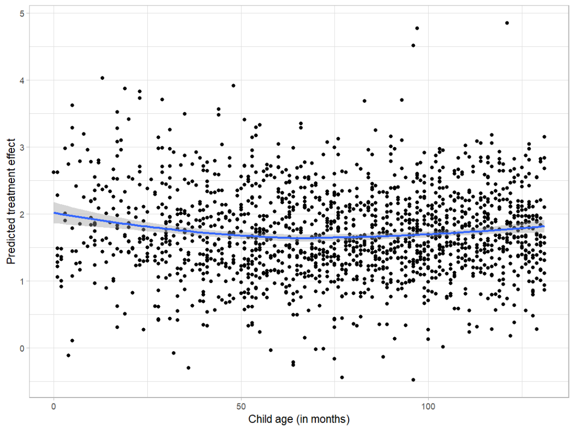
(b) Child health by grandparental age



(c) Maternal satisfaction with the childcare situation by child age



(d) Paternal satisfaction with the childcare situation by child age



Notes: The figures show the predicted treatment effects by child age (panel (a), (c) and (d)) or by grandparental age (panel (b)). The predictions are estimated by using an instrumental forest and the blue line represents the fitted line.
Source: Pairfam (2010-2020), weighted, own calculation.

3.A.4 Robustness Checks

Exclusion of movers. In order to show that our results are not driven by families that (systematically) moved within the observation period, we exclude all households where the distance to grandparents changed from closer/further than 30 minutes to further/closer than 30 minutes. The results for child outcomes are shown in Table 3.A.13 and for parental satisfaction in Table 3.A.14. The coefficients for child health, socio-emotional problems, and school outcomes are similar to the main results. However, the results on child health problems are less statistically significant than the main results. The results for parental satisfaction are similar in magnitude and significance to the main results, which suggests that the results are not driven by (systematic) movement to or away from the grandparents.

Definition of instrument. Furthermore, we check the sensitivity of our results concerning the definition of our instrument. We conduct the analyses with a different binary instrument that equals 1 for all distances shorter than 1 hour away for both *pairfam* and *SOEP*, as well as an ordinal instrument consisting of 6 categories in *pairfam* and 7 categories in the *SOEP*. The results for the alternative binary instrument are presented in Tables 3.A.22 and 3.A.23 and for the ordinal instrument in Tables 3.A.24 and 3.A.25. The results for child outcomes of both alternative specifications are, in terms of magnitude and direction of the effect, quite comparable to our main results. However, the coefficients on child health are statistically less significant. The results with the alternative binary instrument for parental satisfaction are also very similar to our main results. When using the ordinal instrument, the negative effect on paternal satisfaction with education and career is no longer statistically significant, and the effect on maternal satisfaction with the childcare situation is only significant at the 10 percent level.

Grandparental care in hours. In our main specification, we use grandparental care as a binary variable. The *SOEP* data also includes a variable that contains the number of hours a child is cared for by the grandparents. When estimating the effect of grandparental care on children's school outcomes using this variable, we find similar effects to our baseline specification, namely, null effects (Table 3.A.26).

Placebo analysis. Additionally, to further validate our instrument, we estimate the effect of grandparental care on placebo outcomes. We use birth weight (birth weight in grams and a binary variable indicating whether the birth weight is below 2500 grams) for children and the individual's birth month for parents. Both placebo outcomes should not be affected by grandparental care. We do not find any significant effects for either of the outcomes. This supports our empirical approach and the assumption that the method does not show any effects on factors that are independent of grandparental care (see Table 3.A.27).

LPM and "Garden variety". We argue for the use of an LPM model in our main specification as opposed to more conventional non-linear models such as the binary logistic or probit regression models because LPM generates first stage residuals that are uncorrelated with the control variable and fitted values (e.g. Angrist and Pischke, 2008). Furthermore, Hellevik (2009) and Angrist and Pischke (2008) argue that in many applications, LPM generates similar estimates to logit models. To further corroborate our findings and the IV approach, we apply the so-called "garden variety" estimation. In this procedure, one estimates a probit model for the first stage regression and predicts the fitted values after this regression. These non-linear fitted values are then included as an additional instrument in the first stage regression using OLS. The results are presented in Table 3.A.28 and Table 3.A.29. The results on child outcomes are very similar to the main results in terms of magnitude, direction, and significance of the effects. The results on maternal satisfaction with leisure and paternal satisfaction with the childcare situation also match the main results. However, the effect on maternal satisfaction with the childcare situation and paternal satisfaction with education and career become slightly smaller in magnitude and insignificant.

Correction for multiple hypothesis testing. Furthermore, we correct our standard errors for multiple hypothesis testing using the Romano-Wolf Multiple Hypothesis Correction. By doing so, we account for the fact that we conduct a large number of regressions with many different outcomes as testing a large number of hypotheses increases the probability of falsely rejecting a true null hypothesis (Clarke et al., 2020). Applying the Romano-Wolf Correction[37], we obtain a p-value of 0.0640 for maternal satisfaction with leisure, a p-value of 0.0770 for paternal satisfaction with education and career, and a p-value of 0.0730 for child health problems. This means that these effects are statistically significant even when accounting for multiple hypothesis testing.[38]

Further control variables. Finally, we include further control variables to prove the robustness of our results. The results are shown in Tables 3.A.30 and 3.A.31. We include emotional closeness (column 1) and frequency of contact (column 2) as both variables could be related to distance and affect parental satisfaction not only through grandparental childcare. However, since grandparental care could be correlated

[37]We generate 999 bootstrap samples.

[38]As the multiple hypothesis testing command *rwolf* in Stata can only be conducted within one data set, we ran the test for four different groups of outcomes: children's health and socio-emotional outcomes, school outcomes, mother's satisfaction (excluding satisfaction with childcare as it is part of another data set), and father's satisfaction outcomes. Due to the construction of the command, the control variables deviate slightly from our baseline regressions.

to both of these variables, they are potentially bad controls. Therefore, we exclude them from our main set of control variables and include them only in this robustness check. Including these variables does not considerably change the results on either child outcomes or parental satisfaction.

Another factor that might be a threat to the exogeneity assumption is grandparents' health because health limitations have been found to decrease the provision of grandparental care (Hank and Buber, 2009). Additionally, it is plausible that grandparents' illness might have an impact on child outcomes, parents' life satisfaction, and other satisfaction measures. And thirdly, grandparents' health might influence the instrument as families might move closer to a grandparent who is sick and needs help. To prove the robustness of the results, we include two different variables of grandparents' health in our analysis. It can be seen that the inclusion of those variables decreases the sample size considerably. The first variable included in column 3 measures the mean of grandparents' health status during the past 4 weeks. This variable has a lot of missing values because the health status of anchors' parents is surveyed only from wave 2 to wave 7 in the parent questionnaire and not in the anchor and partner questionnaire.[39] Despite the significant decrease in the sample size, the results on the child outcomes change only marginally.[40] However, the results on parental satisfaction become smaller and less significant. In an alternative specification (column 4), we include a variable that indicates whether at least one grandparent needed regular help in the last 12 months and serves as a proxy for bad grandparental health. Although this variable has fewer missing values than the first, it still decreases the sample size considerably. Also, when including this variable, the effects on parental satisfaction decrease and are less significant. In order to find out whether the results actually change because of controlling for grandparental health or whether the sample restrictions due to the many missing values in this variable drive the changes, we conduct the analysis with the restricted sample without controlling for grandparental health. This analysis gives us very similar results to the main results including grandparental health. This suggests that grandparental health does not pose a threat to the exogeneity of our instrument.

We further include the parents' satisfaction value measured before the birth of the first child to account for any individual characteristics that might affect well-being that we haven't accounted for using our instrumental estimator. This reduces the sample size considerably since only households that were part of the survey before the birth

[39] The *pairfam* parent questionnaire is answered by the grandparents. As mentioned in chapter 4, the parent questionnaire is given to anchors' parents if permitted and has a response rate of less than 30 percent (Brüderl et al., 2020).

[40] Note, this analysis is only conducted for the outcomes measured in *pairfam* as this variable is not available in the *SOEP*.

of their first child can be considered. The results in column 5 show that the effects on the mother's satisfaction with leisure are the same size as in our baseline regression. However, the standard error is much larger due to the smaller sample size, which leads to a statistically insignificant coefficient. The negative effect on fathers' satisfaction with career is still found and still significant.[41]

Income and labor force participation are potentially endogenous control variables as they could be correlated with distance and affect our outcome variables not only through grandparental care. Column (5) and (6) in Table 3.A.30 and Column (6) and (7) in Table 3.A.31 show that excluding these variables does not change our estimates and their significance in a substantial way.

[41] This analysis cannot be conducted for satisfaction with the childcare situation because only individuals with children are questioned about their satisfaction with childcare.

Table 3.A.12: Moving behavior before and after the birth of a child

In the year before child birth	General movement	Move towards	Move away from
Any grandparents	0.004	0.003	-0.010
	(0.019)	(0.017)	(0.016)
Observations	22251	22251	22251
Mother's parents	0.018	0.019	-0.003
	(0.017)	(0.013)	(0.012)
Observations	22250	22250	22250
Father's parents	-0.013	-0.016	0.0004
	(0.016)	(0.013)	(0.012)
Observations	20904	20904	20904
In the year after child birth	General movement	Move towards	Move away from
Any grandparents	0.003	0.015	-0.004
	(0.015)	(0.013)	(0.013)
Observations	22251	22251	22251
Mother's parents	0.022	0.010	0.011
	(0.013)	(0.011)	(0.010)
Observations	22250	22250	22250
Father's parents	-0.010	0.006	-0.014
	(0.013)	(0.011)	(0.010)
Observations	20904	20904	20904

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimated using OLS. Standard errors in parentheses. All regressions include individual and household controls described in Table 3.A.2 column (d) except for child-level.

Source: Pairfam (2010-2020), own calculations.

Table 3.A.13: Child outcomes w/o families that moved

	IV: GPC	F-Statistic	Sample Mean	Obs.
Health				
Health problems: 0-2 years	0.224 (0.374)	59.846	1.538	1338
Health problems: 3-5.5 y.	0.276 (0.260)	92.385	1.575	2185
Health problems: 5.5-10 y.	0.258 (0.193)	175.582	1.554	3897
Health problems: 0-10 y.	0.323^{+} (0.191)	192.061	1.563	8289
Socio-emotional behavior				
Socio-emotional problems: 3-5 y.	0.144 (0.284)	79.923	2.936	1596
Conduct problems: 3-5 y.	-0.0610 (0.298)	79.923	1.074	1596
Hyperactivity: 3-5 y.	0.192 (0.260)	80.205	0.987	1597
Emotional problems: 3-5 y.	0.197 (0.256)	80.205	0.874	1597
School outcomes				
Math grade: 9-10 y.	0.0237 (0.186)	79.264	2.251	1420
German grade: 9-10 y.	-0.238 (0.219)	79.463	2.284	1421
Child likes going to school: 9-10 y.	-0.0771 (0.196)	104.135	1.550	2186
Child likes studying: 9-10 y.	0.156 (0.193)	103.941	1.903	2168

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables “the child likes going to schoool” and “the child likes learning” range on a scale from 1 (strongly agree) to 4 (strongly disagree). The original variables on educational aspirations report the probability that a child attains a certain school degree from 1 (not at all) to 7 (completely). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.14: Parental Satisfaction w/o families that moved

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
Mother's Satisfaction with:				
Child care situation	0.843^{+} (0.489)	92.358	8.545	4237
Life	0.408^{+} (0.221)	346.490	7.781	4746
Education, Career	0.364 (0.290)	345.825	7.267	4654
Leisure, Hobbies	0.756^{*} (0.320)	346.564	6.358	4746
Relationship to Partner	0.300 (0.323)	334.311	7.543	4458
Work-life Balance	0.204 (0.417)	190.153	6.462	1933
Father's Satisfaction with:				
Child care situation	1.436^{**} (0.520)	73.370	8.540	3128
Life	0.371^{+} (0.190)	294.787	7.822	3679
Education, Career	-0.528^{*} (0.265)	294.311	7.524	3676
Leisure, Hobbies	-0.018 (0.287)	294.761	6.485	3679
Relationship to Partner	-0.169 (0.360)	294.164	7.706	3676
Work-life Balance	-0.425 (0.400)	193.226	5.953	2059

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.15: Balancing table

	(1)
	Distance
No / lower school degree	-0.0291 (0.0417)
University degree	-0.167*** (0.0301)
Working part-time	0.0116 (0.0212)
Working full-time	0.0351 (0.0300)
Partner working part-time	0.122** (0.0472)
Partner working full-time	0.109** (0.0369)
No partner	0.0763 (0.0555)
Migrational background	-0.0947* (0.0418)
log(income) in 1000€	-0.0504+ (0.0279)
Age	-0.00495 (0.00359)
Children's sex	0.00655 (0.0157)
Nr. children 0-2	0.0981** (0.0311)
Nr. children 3-5.5	0.0967*** (0.0293)
Nr. children 5.5-10	0.0205 (0.0188)
Nr. other children	-0.00201 (0.0182)
Mean GP age	0.000792 (0.00251)
Health	0.000112 (0.00900)
Obesity	-0.0144 (0.0315)
Pregnant	0.0398 (0.0276)
Cohabitation with partner	0.0229 (0.0396)
Widowed	0.108 (0.111)
Single child	0.0725+ (0.0372)
No school degree (grandm.)	0.0787 (0.0630)
Upper school degree (grandm.)	-0.0693+ (0.0371)
No school degree (grandf.)	-0.0781 (0.0758)
Upper school degree (grandf.)	-0.0673+ (0.0359)
Children < 6 in Kita	-0.0695*** (0.0204)
Age youngest child	0.000496 (0.000450)
Children's mean health	-0.0362** (0.0140)
Observations	6395

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimated using OLS based on the sample used in the regressions for maternal satisfaction. Robust standard errors clustered at the household level in parentheses. Regression includes individual and household controls described in Table 3.A.2 column (e).
Source: Pairfam (2009-2020), own calculations.

Table 3.A.16: First stage results (children): Exclusion of controls

Health & Socio-emotional skills:	Health	Socio-emot. problems	Conduct	Hyperactivity	Emotional
Exclusion of controls on parental level					
Distance	0.216***	0.232***	0.233***	0.230***	0.233***
	(0.0150)	(0.0253)	(0.0253)	(0.0254)	(0.0253)
Observations	16839	2741	2743	2745	2743
Exclusion of controls on parental and child level					
Distance	0.220***	0.233***	0.233***	0.231***	0.234***
	(0.0149)	(0.0250)	(0.0250)	(0.0250)	(0.0249)
Observations	16839	2742	2744	2746	2745
Exclusion of controls on (grand-)parental and child level					
Distance	0.198***	0.217***	0.214***	0.216***	0.215***
	(0.0117)	(0.0193)	(0.0193)	(0.0192)	(0.0194)
Observations	26547	5038	5050	5047	5045
Exclusion of all controls					
Distance	0.208***	0.225***	0.224***	0.224***	0.224***
	(0.0117)	(0.0200)	(0.0200)	(0.0199)	(0.0200)
Observations	28426	5363	5376	5374	5370
School outcomes:	Math grade	German grade	Child likes going to school	Child likes studying	
Exclusion of controls on parental level					
Distance	0.293***	0.294***	0.265***	0.265***	
	(0.0377)	(0.0377)	(0.0319)	(0.0321)	
Observations	1498	1499	2309	2293	
Exclusion of controls on parental and child level					
Distance	0.295***	0.296***	0.266***	0.266***	
	(0.0375)	(0.0375)	(0.0317)	(0.0319)	
Observations	1498	1499	2309	2293	
Exclusion of controls on (grand-)parental and child level					
Distance	0.285***	0.285***	0.265***	0.265***	
	(0.0378)	(0.0378)	(0.0318)	(0.0321)	
Observations	1613	1613	2471	2455	
Exclusion of all controls					
Distance	0.277***	0.278***	0.270***	0.269***	
	(0.0432)	(0.0432)	(0.0346)	(0.0349)	
Observations	1663	1663	2538	2522	

Notes: Standard errors in parentheses. Conditional on no missings in the outcome and control variables (see Table 3.A.2).
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.17: First stage results (parents): Exclusion of controls

Parental Satisfaction:	General	Educ./ career	Leisure	Relationship	Work-life balance	Child care
Exclusion of controls on parental level						
Distance: Maternal Sat.	0.225***	0.242***	0.243***	0.242***	0.247***	0.284***
	(0.0181)	(0.00916)	(0.00925)	(0.00914)	(0.00928)	(0.0147)
Observations	9942	11662	11533	11693	11074	4503
Exclusion of controls on parental and child level						
Distance: Maternal Sat.	0.223***	0.241***	0.242***	0.241***	0.245***	0.285***
	(0.0169)	(0.00837)	(0.00844)	(0.00835)	(0.00849)	(0.0145)
Observations	11749	13382	13238	13406	12675	4504
Exclusion of controls on (grand-)parental and child level						
Distance: Maternal Sat.	0.208***	0.227***	0.229***	0.227***	0.234***	0.281***
	(0.0132)	(0.00726)	(0.00732)	(0.00724)	(0.00741)	(0.0131)
Observations	17858	16893	16702	16916	15925	5505
Exclusion of all controls						
Distance: Maternal Sat.	0.204***	0.235***	0.237***	0.235***	0.238***	0.293***
	(0.0128)	(0.00666)	(0.00673)	(0.00665)	(0.00682)	(0.0121)
Observations	19351	18092	17872	18120	17070	5853
Exclusion of controls on parental level						
Distance: Paternal Sat.	0.219***	0.236***	0.236***	0.237***	0.253***	0.246***
	(0.0244)	(0.0107)	(0.0107)	(0.0107)	(0.0106)	(0.0164)
Observations	6857	9685	9677	9690	9376	4722
Exclusion of controls on parental and child level						
Distance: Paternal Sat.	0.208***	0.232***	0.233***	0.232***	0.245***	0.254***
	(0.0228)	(0.00992)	(0.00989)	(0.00992)	(0.00983)	(0.0164)
Observations	8067	11139	11124	11140	10783	4722
Exclusion of controls on (grand-)parental and child level						
Distance: Paternal Sat.	0.188***	0.215***	0.216***	0.215***	0.227***	0.235***
	(0.0176)	(0.00853)	(0.00854)	(0.00854)	(0.00843)	(0.0140)
Observations	12951	13874	13856	13874	13419	5742
Exclusion of all controls						
Distance: Paternal Sat.	0.201***	0.220***	0.222***	0.221***	0.232***	0.239***
	(0.0181)	(0.00797)	(0.00795)	(0.00798)	(0.00793)	(0.0131)
Observations	13756	14770	14760	14780	14284	6082

Notes: Standard errors in parentheses. Conditional on no missings in the outcome and control variables (see Table 3.A.2).
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.18: Child outcomes with entropy balancing

	IV: GPC	F-Statistic	Sample Mean	Obs.
Health				
Health problems: 0-2 years	0.407 (0.295)	102.835	1.546	1828
Health problems: 3-5.5 y.	0.105 (0.203)	176.941	1.579	3006
Health problems: 5.5-10 y.	0.566** (0.186)	195.946	1.573	5132
Health problems: 0-10 y.	0.446** (0.164)	272.789	1.574	11069
Socio-emotional behavior				
Socio-emotional problems: 3-5 y.	0.249 (0.357)	88.023	2.943	2171
Conduct problems: 3-5 y.	0.0661 (0.345)	88.083	1.064	2172
Hyperactivity: 3-5 y.	0.243 (0.299)	88.323	1.002	2173
Emotional problems: 3-5 y.	0.255 (0.339)	88.263	0.878	2172
School outcomes				
Math grade: 9-10 y.	0.128 (0.211)	60.576	2.264	1476
German grade: 9-10 y.	-0.135 (0.244)	60.750	2.300	1477
Child likes going to school: 9-10 y.	-0.164 (0.233)	70.526	1.556	2262
Child likes studying: 9-10 y.	0.205 (0.233)	70.076	1.924	3305

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.

Source: Pairfam (2010-2020), SOEP (2010-2017), entropy balancing weights, own calculation.

Table 3.A.19: Parental Satisfaction with entropy balancing

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
Mother's Satisfaction with:				
Child care situation	1.855** (0.631)	89.448	8.481	5838
Life	-0.0820 (0.253)	237.199	7.759	6182
Education, Career	0.382 (0.351)	232.937	7.171	6061
Leisure, Hobbies	1.038** (0.388)	237.098	6.325	6182
Relationship to Partner	0.288 (0.443)	241.464	7.561	5742
Work-life Balance	0.374 (0.410)	172.977	6.429	2514
Father's Satisfaction with:				
Child care situation	1.810*** (0.503)	68.567	8.496	4011
Life	0.385+ (0.229)	164.924	7.802	4495
Education, Career	-0.277 (0.374)	164.827	7.494	4490
Leisure, Hobbies	0.224 (0.359)	165.277	6 6.451	4494
Relationship to Partner	0.0269 (0.395)	165.496	7.681	4491
Work-life Balance	-0.289 (0.424)	234.580	5.903	2510

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), entropy balancing weights, own calculation.

Table 3.A.20: Parental Satisfaction (using distance to parents-in-law)

Outcomes	IV:GPC	F-Statistic	Sample Mean	Obs.
Mother's Satisfaction with:				
Child care situation	1.558^{+} (0.807)	32.662	8.474	2941
Life	0.207 (0.272)	146.580	7.887	3167
Education, Career	0.273 (0.394)	147.443	7.229	3106
Leisure, Hobbies	0.857^{*} (0.425)	146.318	6.412	3168
Relationship to Partner	0.684^{+} (0.391)	145.586	7.708	3160
Work-life Balance	0.149 (0.540)	85.048	6.553	1284
Father's Satisfaction with:				
Child care situation	1.765^{**} (0.608)	30.243	8.490	2974
Life	0.247 (0.228)	203.211	7.801	3200
Education, Career	-0.024 (0.264)	204.565	7.504	3198
Leisure, Hobbies	0.020 (0.324)	203.986	6.547	3201
Relationship to Partner	0.411 (0.371)	202.544	7.713	3198
Work-life Balance	0.125 (0.448)	136.382	5.954	1787

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.21: Individual Satisfaction (for childless households)

Outcomes	OLS: GPC	Sample Mean	Obs.
Woman's Satisfaction with:			
Life	-0.097 (0.107)	7.747	1266
Education, Career	-0.044 (0.127)	7.364	1265
Leisure, Hobbies	0.108 (0.145)	7.046	1266
Relationship to Partner	-0.050 (0.174)	8.262	1112
Work-life Balance	0.230 (0.254)	6.353	572
Man's Satisfaction with:			
Life	0.060 (0.112)	1120	7.953
Education, Career	-0.131 (0.199)	1117	7.653
Leisure, Hobbies	0.0390 (0.169)	1118	7.061
Relationship to Partner	-0.051 (0.149)	1113	8.273
Work-life Balance	0.296 (0.266)	511	6.080

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimated using OLS. Robust standard errors in parentheses. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Outcomes are on parental level. General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include individual and household controls described in Table 3.A.2 column (e) in the appendix, except for child-level variables.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.22: Results: Child outcomes with different instrument def. (<1h vs. ≥1h)

	IV: GPC	F-Statistic	Sample Mean	Obs.
Health				
Health problems: 0-2 years	0.279 (0.394)	70.558	1.546	1828
Health problems: 3-5.5 y.	0.00881 (0.321)	85.661	1.579	3006
Health problems: 5.5-10 y.	0.298 (0.231)	138.623	1.573	5132
Health problems: 0-10 y.	0.306 (0.223)	168.955	1.574	11069
Socio-emotional behavior				
Socio-emotional problems: 3-5 y.	0.306 (0.341)	45.845	2.943	2171
Conduct problems: 3-5 y.	0.0259 (0.375)	45.742	1.064	2172
Hyperactivity: 3-5 y.	0.212 (0.311)	45.949	1.002	2173
Emotional problems: 3-5 y.	0.473 (0.337)	46.053	0.878	2172
School outcomes				
Math grade: 9-10 y.	-0.429^{+} (0.244)	58.469	2.264	1476
German grade: 9-10 y.	-0.433^{+} (0.256)	58.516	2.300	1477
Child likes going to school: 9-10 y.	0.279 (0.201)	82.248	1.556	2262
Child likes studying: 9-10 y.	0.378^{+} (0.224)	82.973	1.924	3305

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.23: Parental Satisfaction with different instrument def. (<1h vs. ≥1h)

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
Mother's Satisfaction with:				
Child care situation	1.025^{+} (0.622)	82.460	8.481	5838
Life	0.490* (0.248)	275.378	7.759	6182
Education, Career	0.240 (0.350)	272.964	7.171	6061
Leisure, Hobbies	0.792* (0.363)	275.280	6.325	6182
Relationship to Partner	0.454 (0.378)	254.921	7.561	5742
Work-life Balance	0.820 (0.529)	128.672	6.429	2514
Father's Satisfaction with:				
Child care situation	1.690** (0.610)	48.654	8.496	4011
Life	0.177 (0.211)	278.111	7.802	4495
Education, Career	-0.754** (0.273)	277.011	7.494	4490
Leisure, Hobbies	-0.110 (0.337)	278.260	6 6.451	4494
Relationship to Partner	-0.0988 (0.379)	278.809	7.681	4491
Work-life Balance	-1.521** (0.479)	145.997	5.903	2510

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.24: Results: Child outcomes with ordinal instrument definition

	IV: GPC	F-Statistic	Sample Mean	Obs.
Health				
Health problems: 0-2 years	0.464 (0.291)	87.968	1.546	1828
Health problems: 3-5.5 y.	0.157 (0.197)	155.309	1.579	3006
Health problems: 5.5-10 y.	0.240 (0.164)	218.510	1.573	5132
Health problems: 0-10 y.	0.297^{+} (0.152)	264.319	1.574	11069
Socio-emotional behavior				
Socio-emotional problems: 3-5 y.	0.279 (0.231)	97.350	2.943	2171
Conduct problems: 3-5 y.	0.113 (0.241)	97.414	1.064	2172
Hyperactivity: 3-5 y.	0.162 (0.198)	97.559	1.002	2173
Emotional problems: 3-5 y.	0.368 (0.224)	97.495	0.878	2172
School outcomes				
Math grade: 9-10 y.	-0.185 (0.204)	52.365	2.264	1476
German grade: 9-10 y.	-0.175 (0.254)	52.479	2.300	1477
Child likes going to school: 9-10 y.	0.106 (0.189)	76.599	1.556	2262
Child likes studying: 9-10 y.	0.115 (0.183)	76.177	1.924	3305

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.25: Parental Satisfaction with ordinal

Outcomes	IV:GPC	F-Statistic	Sample Mean	Obs.
Mother's Satisfaction with:				
Child care situation	0.809^{+} (0.415)	143.993	8.481	5838
Life	0.237 (0.174)	471.491	7.759	6182
Education, Career	0.162 (0.246)	462.447	7.171	6061
Leisure, Hobbies	0.773^{**} (0.254)	471.494	6.325	6182
Relationship to Partner	0.380 (0.262)	443.013	7.561	5742
Work-life Balance	-0.218 (0.355)	253.774	6.429	2514
Father's Satisfaction with:				
Child care situation	1.504^{***} (0.433)	74.011	8.496	4011
Life	0.350^{*} (0.171)	355.295	7.802	4495
Education, Career	-0.126 (0.225)	354.859	7.494	4490
Leisure, Hobbies	-0.166 (0.264)	356.177	6 6.451	4494
Relationship to Partner	-0.0182 (0.313)	355.613	7.681	4491
Work-life Balance	-0.274 (0.350)	216.907	5.903	2510

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Child care", robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome "Child care" and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.26: Results: Child outcomes with linear grandparental care variable

	IV: GPC	F-Statistic	Sample Mean	Obs.
School outcomes				
Math grade: 9-10 y.	0.00639	26.606	2.264	1475
	(0.0281)			
German grade: 9-10 y.	-0.0148	26.789	2.300	1476
	(0.0302)			
Child likes going to school: 9-10 y.	-0.00984	30.944	1.556	2278
	(0.0216)			
Child likes studying: 9-10 y.	0.0237	30.662	1.924	2261
	(0.0241)			

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 (c) for school outcomes in the appendix.
Source: SOEP (2010-2017), weighted, own calculation.

Table 3.A.27: Placebo Regressions

Outcomes	IV:Grandparental Care	F-Statistic	Sample Mean	Obs.
Child:				
Birth weight	131.8 (259.9)	115.776	3420.660	6606
Birth weight < 2500	-0,0571 (0.0976)	115.776	0.049	6606
Parents:				
Mother: Birth month	-0.436 (0.492)	328.651	6.653	6183
Father: Birth month	-0.860 (0.597)	219.983	6.459	4485

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome "Birth weight", robust standard errors clustered at the household level. The regressions include the control variables listed in Table 3.A.2 column (a) for the outcomes on birth weight and (e) for the outcomes on birth month in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.28: Results: Child outcomes (applying "Garden Variety")

	IV: GPC	F-Statistic	Sample Mean	Obs.
Health				
Health problems: 0-2 years	0.532^{+} (0.292)	39.371	1.551	1811
Health problems: 3-5.5 y.	0.455^{*} (0.217)	63.068	1.585	2990
Health problems: 5.5-10 y.	0.387^{*} (0.173)	88.440	1.587	5116
Health problems: 0-10 y.	0.496^{**} (0.163)	110.256	1.584	11040
Socio-emotional behavior				
Socio-emotional problems: 3-5 y.	0.0900 (0.231)	44.001	3.013	2164
Conduct problems: 3-5 y.	-0.0440 (0.258)	44.084	1.092	2165
Hyperactivity: 3-5 y.	0.0222 (0.223)	44.241	1.018	2166
Emotional problems: 3-5 y.	0.232 (0.238)	44.158	0.903	2165
School outcomes				
Math grade: 9-10 y.	0.0378 (0.186)	39.197	2.264	1476
German grade: 9-10 y.	-0.0897 (0.220)	39.297	2.300	1477
Child likes going to school: 9-10 y.	-0.0571 (0.205)	50.449	1.556	2262
Child likes studying: 9-10 y.	0.119 (0.197)	50.461	1.924	2245

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The original variables on educational aspirations report the probability that a child attains a certain school degree from 1 (not at all) to 7 (completely). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.29: Parental Satisfaction (applying “Garden Variety”)

Outcomes	IV: GPC	F-Statistic	Sample Mean	Obs.
Mother’s Satisfaction with:				
Child care situation	0.581 (0.420)	58.574	8.471	5834
Life	0.0294 (0.173)	212.183	7.744	6182
Education, Career	0.412^{+} (0.237)	207.308	7.163	6061
Leisure, Hobbies	1.057^{***} (0.268)	212.103	6.322	6182
Relationship to Partner	0.114 (0.251)	214.140	7.560	5742
Work-life Balance	0.0656 (0.343)	129.352	6.406	2514
Father’s Satisfaction with:				
Child care situation	1.651^{***} (0.443)	41.440	8.476	3980
Life	0.157 (0.166)	169.167	7.798	4495
Education, Career	-0.229 (0.221)	168.942	7.484	4490
Leisure, Hobbies	-0.315 (0.264)	169.172	6.465	4494
Relationship to Partner	-0.266 (0.295)	168.800	7.691	4491
Work-life Balance	-0.340 (0.376)	116.856	5.919	2510

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome “Child care”, robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome “Child care” and (e) for all other outcomes in the appendix.
Source: Pairfam (2010-2020), weighted, own calculation.

Table 3.A.30: Results: Child outcomes (including/excluding controls)

Outcomes *including/excluding*	Emot. Closeness	IV:Grandparental Care Freq. Contact	GP Health	GPH (Proxy)	Excl. Income	Excl. LFP
Health						
Health prob.: 0-2 years	0.580 (0.384)	0.712 (0.490)	0.244 (0.373)	0.667 (0.413)	0.525 (0.342)	0.548 (0.360)
Observations	1828	1828	446	663	1936	1869
Health prob.: 3-5.5 y.	0.323 (0.241)	0.438 (0.297)	0.516^{+} (0.294)	0.528^{+} (0.271)	0.138 (0.252)	0.244 (0.234)
Observations	3006	3006	628	1233	3198	3104
Health prob.: 5.5-10 y.	0.493^{*} (0.199)	0.615^{*} (0.248)	0.572^{*} (0.281)	0.609^{*} (0.242)	0.464^{*} (0.199)	0.461^{*} (0.194)
Observations	5132	5132	851	1874	5452	5311
Health prob.: 0-10 y.	0.530^{**} (0.190)	0.669^{**} (0.238)	0.508^{*} (0.227)	0.616^{**} (0.195)	0.462^{*} (0.186)	0.486^{**} (0.182)
Observations	11069	11069	2091	4130	11772	11450
Socio-emotional behavior						
Socio-emot. prob.: 3-5 y.	0.431 (0.298)	0.533 (0.395)	0.225 (0.336)	0.443 (0.302)	0.381 (0.285)	0.418 (0.278)
Observations	2171	2171	474	742	2286	2241
Conduct prob.: 3-5 y.	0.273 (0.318)	0.289 (0.410)	0.310 (0.319)	0.394 (0.333)	0.119 (0.308)	0.288 (0.308)
Observations	2172	2172	474	742	2287	2242
Hyperactivity: 3-5 y.	0.333 (0.268)	0.498 (0.373)	0.233 (0.331)	0.192 (0.290)	0.348 (0.269)	0.239 (0.259)
Observations	2173	2173	474	742	2288	2243
Emotional prob.: 3-5 y.	0.363 (0.297)	0.398 (0.382)	-0.0559 (0.360)	0.424 (0.321)	0.395 (0.289)	0.420 (0.282)
Observations	2172	2172	474	742	2287	2242
School outcomes						
Math grade: 9-10 y.					0.0537 (0.199)	0.0770 (0.211)
Observations					1522	1487
German grade: 9-10 y.					-0.120 (0.230)	-0.170 (0.243)
Observations					1523	1488
Likes school: 9-10 y.					-0.0435 (0.219)	-0.118 (0.223)
Observations					2342	2295
Likes studying: 9-10 y.					0.250 (0.210)	0.213 (0.219)
Observations					2325	2278

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors clustered at the household level in parentheses. All outcome variables are standardized with mean 0 and standard deviation of 1. The original general health problems variable is an ordinal variable on a scale from 1 (good health) to 5 (bad health). The original outcome variables conduct problems, hyperactivity and emotional problems are ordinal variables on a scale from 0 (does not apply) to 5 (fully applies). The outcome variable socio-emotional problems is constructed summing up the three other indices, resulting in a variable that ranges from 0 (does not apply) to 12 (fully applies). The original outcome variables for math and German grade measure the school grades in these two subjects from 1 (very good) to 6 (very bad). The original variables "the child likes going to schoool" and "the child likes learning" range on a scale from 1 (strongly agree) to 4 (strongly disagree). The regressions include the control variables listed in Table 3.A.2 column (a) for health problems, (b) for socio-emotional problems and (c) for school outcomes in the appendix. LFP = Maternal labor force participation, GPH = Grandparental health.
Source: Pairfam (2010-2020), SOEP (2010-2017), weighted, own calculation.

Table 3.A.31: Parental Satisfaction (including/excluding controls)

Outcomes	IV:Grandparental Care						
including/excluding	Emot. Closeness	Freq. Contact	GP Health	GPH (Proxy)	Pre-birth sat.	Excl. Income	Excl. LFP
Mother's Sat.:							
Child care	0.843^{+} (0.473)	0.949 (0.605)	1.338^{+} (0.707)	0.0862 (0.599)		1.011^{*} (0.445)	0.898^{+} (0.460)
Observations	5838	5838	1120	2224		6289	5850
Life	-0.136 (0.216)	-0.234 (0.255)	-0.121 (0.404)	0.152 (0.335)	-0.170 (0.420)	0.106 (0.214)	0.0413 (0.209)
Observations	6174	6182	1053	2039	1903	6563	6191
Educ., Career	0.268 (0.299)	0.146 (0.347)	-0.422 (0.570)	0.270 (0.452)	1.347^{*} (0.601)	0.496^{+} (0.297)	0.468 (0.294)
Observations	6053	6061	1043	1996	1845	6429	6070
Leisure	0.785^{*} (0.317)	0.896^{*} (0.372)	-0.263 (0.613)	0.528 (0.461)	0.818 (0.647)	1.175^{***} (0.315)	0.844^{**} (0.305)
Observations	6174	6182	1053	2039	1901	6564	6191
Relationship	0.139 (0.320)	0.249 (0.366)	-0.296 (0.559)	0.217 (0.500)	0.597 (0.503)	0.249 (0.315)	0.213 (0.310)
Observations	5736	5742	990	1892	1727	6111	5748
Work-life Bal.	0.0667 (0.394)	0.145 (0.456)	-0.879 (0.671)	-0.451 (0.626)	4.494 (3.470)	0.162 (0.376)	0.0596 (0.387)
Observations	2512	2514	348	900	156	2650	2514
Father's Sat.:							
Child care	1.701^{**} (0.554)	1.709^{*} (0.682)	0.383 (0.823)	1.779^{**} (0.606)		1.752^{**} (0.536)	1.852^{**} (0.566)
Observations	4011	4011	716	1532		4194	4342
Life	0.0491 (0.210)	-0.102 (0.262)	-0.246 (0.305)	-0.213 (0.313)	-0.0631 (0.306)	0.0908 (0.200)	0.206 (0.204)
Observations	4011	4495	664	1464	1733	4721	4499
Educ., Career	-0.674^{*} (0.284)	-1.030^{**} (0.359)	-0.902^{*} (0.414)	-0.110 (0.411)	-0.684^{+} (0.399)	-0.650^{*} (0.272)	-0.509^{+} (0.277)
Observations	4488	4490	664	1463	1726	4715	4494
Leisure	-0.119 (0.330)	-0.183 (0.408)	-0.426 (0.512)	0.116 (0.478)	-0.635 (0.492)	-0.0208 (0.317)	-0.0455 (0.318)
Observations	4492	4494	664	1463	1727	4719	4498
Relationship	-0.432 (0.364)	-0.643 (0.449)	0.310 (0.533)	-0.460 (0.576)	-0.715 (0.553)	-0.248 (0.351)	-0.236 (0.354)
Observations	4489	4491	663	1461	1672	4716	4495
Work-life Bal.	-0.412 (0.435)	-0.751 (0.553)	-1.363^{+} (0.794)	-0.858 (0.658)	2.374^{*} (1.191)	-0.228 (0.412)	-0.387 (0.426)
Observations	2509	2510	316	880	369	2632	2512

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. For the outcome “Child care”, robust standard errors clustered at the household level. The outcome variables are all ordinal variables on a scale from 0 (very dissatisfied) to 10 (very satisfied). Child care: satisfaction with the child care situation (on child level, all other outcomes on parental level), General: general life satisfaction, Educ./career: Satisfaction with education and career, Leisure: satisfaction with leisure and hobbies, Relationship: satisfaction with the relationship with the current partner, Work-life balance: satisfaction with the proportion of time that individuals spend on the job or for vocational training or university education relative to the time that individuals spend on personal life. The regressions include the control variables listed in Table 3.A.2 column (d) for the outcome “Child care” and (e) for all other outcomes in the appendix. LFP = Maternal labor force participation, GPH = Grandparental health.
Source: Pairfam (2010-2020), weighted, own calculation.

CHAPTER 4

Age-specific Effects of Early Daycare on Children's Health[1]

4.1 Introduction

Since the early 2000s, the share of very young children (0–2 years) in daycare has increased significantly in many OECD countries.[2] Germany experienced one of the largest increases among all OECD countries from a 17% coverage rate in 2005 to 37% in 2018 (OECD, 2019a). Along with this development, the body of literature studying the effects of (early) daycare attendance of children on their socio-emotional and cognitive outcomes has grown. Previous research shows that health is one of the most important determinants of socio-emotional and cognitive development during childhood, and of later educational achievements, health outcomes, and labor market outcomes during adulthood (e.g., Carneiro et al., 2007; Currie, 2020; Currie and Stabile, 2006; Heckman et al., 2013; Heckman, 2007; Peet et al., 2015). Additionally, child health *per se* matters – also because ill health produces direct costs to the healthcare system as well as indirect societal costs through, for example, labor productivity losses of parents. However, despite the relevance, the effect of early daycare attendance on health receives little attention in the literature.

[1]I am grateful to the National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung, KBV) for data access and for their excellent support. I also thank C. Katharina Spieß, Peter Haan, Clara Schäper, Laura Schmitz, Elena Ziege, Jan Marcus, Mathias Huebener, Katharina Wrohlich, Virginia Sondergeld, Dominik Sachs, Jonas Jessen and Hannes Schwandt for helpful comments and fruitful discussions on the topic. Finally, my thanks goes to the participants in the CRC Summer School 2021: Science-Based Policy Advice, Berlin, the BeNA Summer Workshop 2021, Berlin, the LEER Conference on Education Economics 2022, Leuven, the Spring Meeting of Young Economists (SMYE) 2022, Orléans, the Annual meeting of the Society of Economics of the Household (SEHO) 2022, London, the EEA-ESEM (European Economic Association) 2022, Milan, the 1st Berlin Workshop on Empirical Public Economics 2022, Berlin and internal seminars at DIW Berlin.

[2]On average, the enrollment rate of 0–2-year-old children in daycare increased from 21% in 2005 to 32% in 2018 (e.g., OECD, 2019a).

The few studies that assess the effects of daycare exposure on child health are inconclusive in terms of the direction and magnitude of the effects. Programs targeting families from low socio-economic backgrounds generally benefit children's health (e.g., Conti et al., 2016; Hong et al., 2019) while the effects of universal daycare programs on health depend on the quality of the program[3], the counterfactual care mode, the considered age groups and the outcomes at measure. For example, a cheap, low-quality daycare expansion in Quebec has adverse health effects (e.g., Baker et al., 2008, 2019; Heckman et al., 2010; Kottelenberg and Lehrer, 2013), while daycare expansions providing better quality care have the potential to have null or positive effects on child health (e.g., Bosque-Mercader, 2022; Cornelissen et al., 2018; van den Berg and Siflinger, 2022).[4] Evidence on the health effects of early daycare entry (below age three) is particularly scarce, as most studies either focus on older children or on daycare expansions that affect all children below school age.

In this paper, I fill this research gap by analyzing the age-specific effects of early daycare[5] attendance of children on their mental and physical health. To overcome the endogeneity of the decision to attend daycare at an early age, I exploit a large-scale expansion of publicly funded daycare in Germany. This expansion was induced by a federal reform that introduced a legal entitlement to 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 by about 17 percentage points between 2008 and 2018 in West Germany (Destatis, 2019a). 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 to identify causal effects.[6]

The analyses are based on administrative health records covering all individuals insured through the public health system in Germany (about 90% of the population) between 2009 and 2019. My sample includes children from birth cohorts 1999 to 2015 aged one to ten years, 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

[3]There is evidence that the quality of daycare is at least equally important for child development, including health, as daycare attendance *per se* (see, e.g., Blanden et al., 2022; Kuger et al., 2019; Peter, 2013; Spiess, 2022, and references therein).

[4]For a literature review on the impact of universal early education programs, especially on health, see, e.g., Cascio (2015), Dietrichson et al. (2020), van Huizen and Plantenga (2018), and, with a focus on Germany, Spiess (2022).

[5]The term daycare describes all forms of formal childcare provided by professionals outside the family.

[6]The expansion allows a clear treatment definition that does not require applying DiD estimators developed for staggered treatment implementations (e.g., Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020).

by practitioners based on the International Classification of Diseases (ICD-10) are recorded for each visit. Specifically, I consider physical and mental health outcomes, healthcare consumption, and costs. In terms of physical health, I analyze three sets of communicable diseases – *infections*, *respiratory diseases*, and *ear diseases* – and three non-communicable diseases – *obesity*, *injuries* and *vision problems*. For mental health, including socio-emotional outcomes, I consider the ICD-10 group of *mental and behavioral disorders*. To measure healthcare consumption, I assess the annual number of treatment cases and healthcare costs. *Ex-ante*, there is no clear prediction for the direction of the effects as the daycare expansion may affect these outcomes through several channels. Underlying channels include the earlier onset of an immunization process, formation of health habits, formation of socio-emotional and cognitive skills, and changes in the child's environment other than daycare attendance per se (e.g., health surveillance by the daycare teachers, increased maternal labor market participation and improved parental well-being).

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. Specifically, a ten percentage point increase in the daycare coverage rate leads to an increase of 0.08 infections, 0.03 ear diseases, and 0.016 respiratory diseases per child per year at age one to two. These estimates correspond to a 5.7% increase for infections, 5.1% for ear diseases, and 5.6% for respiratory diseases compared to the sample means. The reductions in infections and respiratory diseases at elementary school age are of similar magnitude in absolute terms. In line with the hygiene hypothesis, which states 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),[7] my results suggest a substitution of illness spells from elementary school to the first years of daycare. The increases in infections and respiratory diseases at age 1–2 years correspond to the decreases at elementary school age, suggesting that children who enter daycare earlier suffer from the same number of infections and respiratory diseases during their first ten years of life as children who enter daycare later. I do not find robust evidence of significant changes in mental health or obesity, while my results suggest null effects on injuries and vision problems. Healthcare consumption increases at ages 1–2 while it decreases at ages 3–5 and 6–8. Despite changes in the prevalence of diagnoses and the number of doctor visits, there is

[7] 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. Subsequently, the hygiene hypothesis was also 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).

no clear effect on healthcare costs. The findings are robust to a large set of robustness checks, such as different definitions of the treatment status and the expansion period, the application of multiple hypothesis testing methods to obtain p-values accounting for the large number of outcomes, and plausibility checks of the common trend assumption. Heterogeneity analysis indicates more pronounced effects for children from disadvantaged areas, earlier detection of vision problems and a reduction in obesity in these children.

These results raise the question of whether substituting illness spells for infections and respiratory diseases from elementary school to the first years of daycare is beneficial. The daycare expansion appears to be neutral in terms of healthcare costs arising in the first ten years of life. My results suggest that the beneficial health effects for older children may reach beyond the study period, which aligns with previous literature on long-term daycare effects. I provide suggestive evidence from an additional analysis based on representative survey data from the German Socio-Economic Panel (SOEP) for spill-over effects on parents, i.e., that parental health deteriorates in the short run but improves in the long run. Furthermore, sickness absence at work of mothers with elementary school-age children is lower when their children entered daycare at age 1–2 years. This, in turn, may increase productivity as mothers tend to work more hours when children are in elementary school compared to when children are below three years (Federal Institute for Population Research, 2020). In terms of spill-over effects on siblings, evidence from Daysal et al. (2022) points out that older siblings "bringing home" infections from daycare leads to worse health for younger siblings, who are particularly vulnerable below the age of one. Thus, moving infections to an earlier age when there are no younger siblings could benefit future younger siblings. When classifying the results, I also reflect on other factors such as the duration of illness spells at different ages and sickness absence at school or daycare. Overall, there is no evidence that changing the timing of infections to earlier years leads to detrimental effects that would challenge children's daycare entry at an early age.

My study contributes to the literature in several ways. First, this is the first study to specifically estimate the health effects of a universal daycare program for children below three.[8] Van den Berg and Siflinger (2022), who assess a reform that abolished daycare fees for children aged one to five years in one region in Sweden, evaluate the cumulative effect of attending daycare from age one until school entry. In Germany, prior to the reform, daycare attendance from age three was already almost universal; thus, the reform only shifted the daycare entry age to an earlier age. Second, I estimate age-specific effects by assessing instantaneous effects on child health (age one to two)

[8] An exception is Lauber (2015), who studies the effects of daycare enrollment at 30 months on obesity.

as well as short-term effects (age three to five) and longer-term effects at elementary school age (age six to ten). Assessing age-specific effects is important as the effect of daycare may change over the life-course (e.g., Cattan et al., 2021; van den Berg and Siflinger, 2022).[9]

Third, my detailed diagnosis data enables me to understand the potential heterogeneity of the health effects. Most previous studies rely on survey data that contain rather broad and subjective health measures. Survey data allow for assessing health and behavioral outcomes that cannot be measured otherwise and usually provide an extensive range of socio-economic characteristics. However, when measuring health and behavioral disorders, survey data are less detailed than administrative health records and are potentially subject to a reporting bias (e.g., Bound et al., 2001). In order to obtain a comprehensive understanding of the effects of daycare on health, it is essential to study various dimensions as different diseases could be differently affected. So far, there are only two studies assessing daycare's health effects using detailed administrative health records; namely, van den Berg and Siflinger (2022) and Bosque-Mercader (2022).

Lastly, I study parental care as the counterfactual (i.e., alternative) care option. In Germany, prior to the expansion, there were almost no care options outside the family[10] and maternal labor market participation was low.[11] In other institutional contexts and countries the counterfactual care mode is different. For example, van den Berg and Siflinger (2022) study a reform that abolished daycare fees in one region in Sweden, which led to a switch from non-parental care to formal daycare arrangements. Thus, the Swedish reform led to a less drastic change for the children than the German reform, which induced a move from (grand)parental care to daycare.

This paper is structured as follows. In the next section, I discuss the channels through which daycare affects child health in detail. After that, I describe the institutional setting, particularly the daycare expansion. In section 4.4, I present my data focusing 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 4.6, I present my empirical results, discuss them and provide a heterogeneity analysis as well as an extensive set of robustness checks. Section 4.7 concludes.

[9] For an overview see, e.g., Burger (2010).

[10] The main care actors of children below three are parents but also about 30% of children are cared for by other relatives, mainly grandparents (Barschkett et al., 2021b).

[11] The daycare expansion in Germany is shown to increase labor market participation of mothers with young children (Müller and Wrohlich, 2020).

4.2 Potential mechanisms

Since daycare attendance may affect health through several channels, it is difficult to anticipate the direction of the effects. First, communicable diseases such as infections, respiratory conditions, and ear problems are very prevalent among (young) children, which is in line with the hygiene hypothesis (Strachan, 1989). In fact, the results of van den Berg and Siflinger (2022), alongside evidence from the medical and epidemiological literature, suggest 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).[12] Specifically, van den Berg and Siflinger (2022) use detailed administrative health records for one region in Sweden and exploit a daycare reform that increased daycare exposure by reducing fees for public daycare. Their results suggest that daycare attendance improves mental health at elementary school age and substitutes infections from elementary school ages to younger ages. Non-parental care arrangements serve as the counterfactual to public daycare in this setting.[13][14] Similarly, injuries are likely to happen more frequently when children interact with other children. For example, Barschkett et al. (2021a) find that during Covid-19 lockdowns, when daycare centers were closed, the number of diagnoses for injuries reduced.[15] Furthermore, the expansion reduced child maltreatment (Sandner and Thomsen, 2020), which could be reflected in diagnosed injuries. Additionally, evidence from the Sure Start program – an early education program in the UK offering a range of services to support children and parents – supports the arguments on infectious diseases and injuries. Cattan et al. (2021) evaluate short- and medium-term health impacts using administrative health records on hospital admissions. They find exposure to Sure Start leads to an increase in hospitalizations at age one and a decrease at age 11–15. The main drivers of the increase in hospitalizations are infectious diseases in the short run, while in the long run, admissions due to accidents and injuries, infectious illnesses, and mental health-related conditions decrease.

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

[13] Lundin et al. (2008) show that the reduction in daycare fees had no effect on the labor market participation of Swedish women. This suggests that prior to the reform, children were already in non-parental care.

[14] Aalto et al. (2019) study the health effects of daycare on children with unemployed parents. 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 due to reduced prescriptions for allergies and asthma at elementary school age.

[15] Note, this effect is a general "lockdown"-effect and cannot be interpreted causally as a daycare effect.

Second, daycare teachers may play an essential 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. Obesity has its onset often early in childhood[16] and is influenced by health behavior and general lifestyle. Being overweight or obese is an important determinant of skill development during childhood (e.g., Cawley and Spiess, 2008) and of future health problems and chronic conditions (such as cardiovascular diseases and diabetes, Must et al., 1999). Interventions to prevent obesity are shown to be particularly effective in children younger than six (Davis and Christoffel, 1994; Waters et al., 2011). Additionally, eating habits – which are crucial causes of obesity – are likely developed early in life (e.g., Birch, 1999). Therefore, daycare attendance can influence health habits and, as a result, prevent obesity. Evidence from a universal daycare expansion during the 1990s for children three and older in Germany supports this argument. Specifically, the reform was shown to positively affect children's physical health, i.e., a decline in physician recommendations for compensatory sport (Cornelissen et al., 2018) as well as fewer weight problems and better performance in the gross motor skills test (Lauber, 2015).

Third, there is evidence that daycare attendance is associated with the development of socio-emotional skills (e.g., Baker et al., 2008; Felfe and Lalive, 2018; Peter et al., 2016) and that the formation of socio-emotional skills is at least equally as important as the development of cognitive skills.[17] For example, Currie and Stabile (2006) point out that mental disorders have larger adverse effects on future reading and mathematics test scores than physical health problems. Furthermore, there is evidence that daycare can affect the salivary cortisol level in young children. Higher cortisol levels can be evidence of stress and decrease the antibody levels, which can result in greater illness frequency (Watamura et al., 2010).

Lastly, not only daycare attendance *per se* but also changes in the child's environment due to the expansion may affect their health. For example, some kind of health surveillance at daycare centers might track children's health (e.g., traces of abuse, detection of vision problems). Attending daycare most likely does not affect the likelihood of having eye problems but rather the *timing* of detecting such problems. Hong et al. (2019) provide evidence that attending a pre-kindergarten program in the US increases the probability of being diagnosed with vision problems, thus leading to earlier onset of treatment. Additionally, as Müller and Wrohlich (2020) show, the daycare expansion increased female labor market participation. On the one hand, employed parents need

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

[17] There is a large body of literature assessing the effects of daycare on cognitive as well as socio-emotional (mostly measured by child development inidices such as the SDQ) skills. For an overview on this literature, see, e.g., Baker (2011); Elango et al. (2015).

a doctor's note to take sick leave when their child is sick. On the other hand, employed parents have more time pressure than parents who are not employed. Thus, the incentives to take their child to the doctor more or less often for employed parents could go either way. Furthermore, Schmitz (2020) provides evidence that daycare attendance of children can have positive effects on parental well-being which in turn positively influences children (Berger and Spiess, 2011; Coneus and Spiess, 2012a).[18] Based on these contradicting predictions, it remains an empirical question whether the expansion affected children's health outcomes and whether they improved or deteriorated in specific age groups.

One must further note that the effects depend on the counterfactual care mode. In Germany, the counterfactual care mode is family care (mainly provided by parents but also grandparents), while, for example, in Scandinavia, the counterfactual care mode is mostly non-parental care arrangements. Therefore, the differences in the children's environment when moving from family care to daycare is probably more significant than moving from non-parental care to daycare.

4.3 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).[19] Besides incentives set by the tax and transfer system, one frequently quoted reason is the low supply of daycare for (young) children. Since 1996, every child aged three and older has been legally entitled for a daycare slot. As of 2022, almost all children visit a daycare center for at least one year before entering school (Destatis, 2022a). Other policy reforms affecting the supply of all-day slots and daycare slots for children younger than three have only been initiated since the middle of the 2000s (Spieß, 2011). In 2005, the daycare expansion law ("Tagesbetreuungsausbaugesetz", TAG) was passed, aiming to expand daycare slots for children under the age of three (230,0000 additional slots in West Germany). In 2007, a summit of the federal government, the federal states, and the counties reinforced the aim of the 2002 EU-mandate[20] and set the target of a 35% daycare coverage

[18]For a recent literature overview on the effects of daycare on various dimensions, including maternal labor market participation and child outcomes, with a focus on Germany, see Spiess (2022).

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

[20]In 2002, the European Council set objectives regarding the provision of daycare in the "Barcelona objectives" (European Council, 2002). By 2013, all member states should provide daycare for at least 33% of children below three.

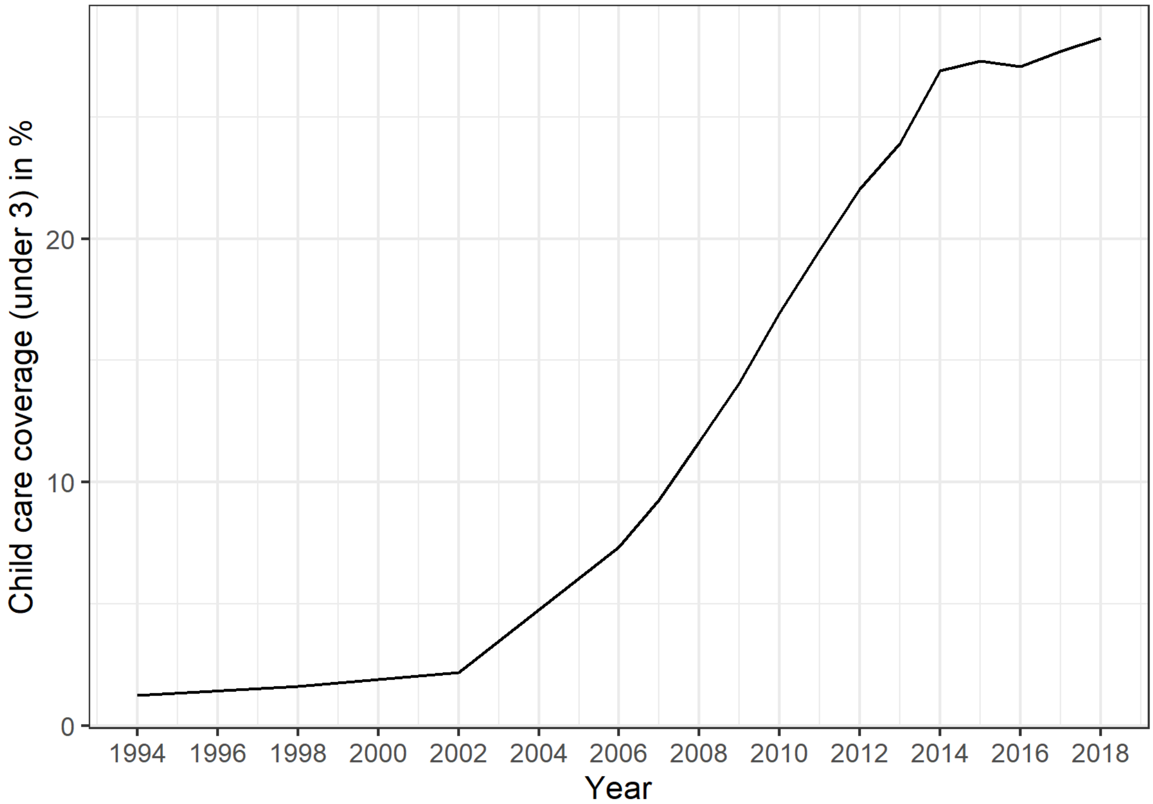
rate for children under three years by 2013. Finally, in 2008 the law on support for children ("Kinderförderungsgesetz") was introduced, committing states to a gradual expansion of daycare supply for children below three years. The law also entailed a legal entitlement to every parent with a child aged one to three years to a subsidized daycare slot (in a daycare center or with a childminder) 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% over the same period (Destatis, 2018).[21] I restrict the analysis to West Germany (excluding Berlin) as the situation in West and East Germany is not comparable and the expansion was significantly larger in West Germany. The development of daycare coverage in West Germany between 1994 and 2018 is depicted in Figure 4.1. It becomes visible that daycare coverage for children under three years was very low (below 5%) up until the early 2000s. From the mid-2000s, West Germany experienced a steep increase that ran flatter from 2014 onward. The increase experienced until 2018 was significant; however, the goal of a 35% coverage was not reached. Furthermore, the expansion created sizeable regional variation in the expansion speed. Figure 4.2 shows that, in 2008, the majority of West German counties had a coverage rate below 20%. In 2018, the majority of counties lay above 20%, many above 30%. Additionally, it becomes apparent that the expansion speed differed substantially across counties.

Germany is characterized by a publicly subsidized daycare system. One third of publicly funded daycare slots is provided by local authorities or municipalities, while private providers that are mostly publicly subsidized account for the remaining slots. Private providers include religious non-profit (one third), non-religious non-profit (17 percent), and other providers (15 percent) (Muehler, 2010; Spiess, 2022). Overall, 98% of providers are considered non-profit providers (Destatis, 2018). Generally, daycare is highly subsidized by the federal government, the states, and the municipalities, but the exact amount and source of funds varies by state. Compared to other OECD countries, Germany's public expenses relative to the GDP are slightly above average. Daycare fees typically range between 5 to 9% of net family income (Schmitz et al., 2017), which

[21] Daycare coverage is defined as the share of children being in daycare, entailing daycare centers and childminders. The majority of children visit a daycare center. In 2018, only 5.4% were cared for by a childminder (Destatis, 2018). As there is ongoing, persistent, 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).

Figure 4.1: Daycare coverage (under 3) West Germany



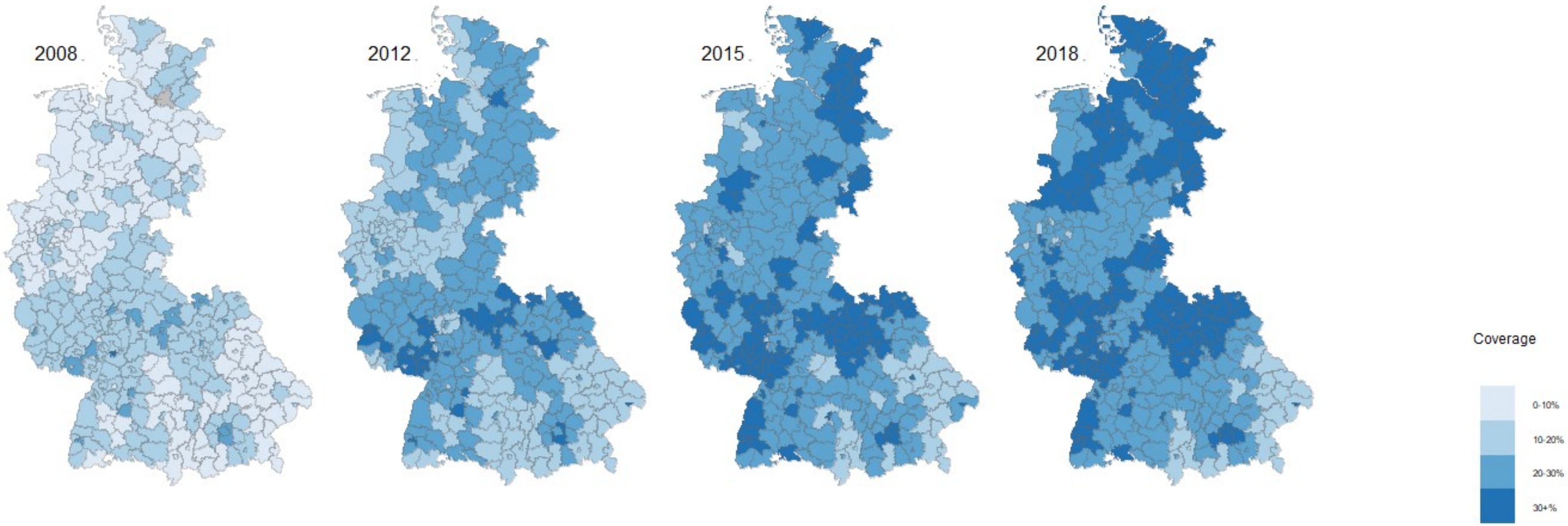
Notes: The graph shows the daycare coverage rate for children below three years in West Germany.
Source: Destatis 1994-2018, own calculations.

is below the OECD average (OECD, 2019a).[22] Most of the general objectives, strategies, and funding sources of daycare are determined at the federal level. However, operational planning and the implementation of objectives are managed by municipal governments and/or youth welfare offices. Thus, the structure and organization of daycare vary between states and communities (Hüsken, 2011; Müller and Wrohlich, 2020). Local authorities estimate the local demand for daycare slots and develop an expansion strategy accordingly. However, the procedure is not uniform across municipalities, thus leading to the observed differences in the expansion speed.

Previous work has established that the expansion increased female labor market participation (Müller and Wrohlich, 2020), increased fertility (Bauernschuster et al., 2016), reduced child maltreatment (Sandner and Thomsen, 2020), and improved children's socio-emotional skills Felfe and Lalive (2018).[23] Furthermore, even though daycare is universal and open to all families, the take-up rate of the scarce daycare slots for chil-

[22]For a more detailed overview of the organization and funding of German daycare centers, see, e.g., Huebener et al. (2020).
[23]For an overview of the effects of the daycare expansion, see, e.g., Rainer et al. (2013); Spiess (2022).

Figure 4.2: Daycare expansion during 2008 and 2018



Notes: 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.

dren below the age of three is higher among highly educated and non-migrant families (e.g., Jessen et al., 2020).

4.4 Data

For the analysis, I use administrative data covering 2009 through 2019,[24] collected by all public health insurers 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 public and private insurance systems. Nearly 90% of the German population is covered by one of the public health insurance funds.[25] Only individuals with earnings exceeding a certain threshold[26] and individuals in specific occupational groups (e.g., civil servants and self-employed) are allowed to opt out of the public system and sign up with a private insurance company instead. The health insurance of their parents covers children without extra fees.

4.4.1 Sample

I have access to data covering 2009 through 2019; thus, I am able to 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 4.3, I construct a treatment and a control group depending on the children's birth year and their county of residence when I first observe them. In the analysis, I focus on children born between 1999–2015. Comparing the number of public health insured children (BMG, 2020) with

[24]Data are only available from 2009, and years from 2020 are excluded from the analysis due to the Covid-19 pandemic.

[25]Mandatory contributions from employers and employees combined with tax revenues are the primary financing sources of the German public health insurance. First, contributions are pooled in a Central Health Fund. Secondly, the contributions are reallocated to the sickness funds according to a morbidty-based risk adjustment scheme. For more information about the German health insurance system, see OECD (2019b).

[26]The income threshold for 2022 was 64,350 euro ($\approx 62,734$ dollar) per year.

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.[27][28]

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 period. Thus, the sample is unbalanced because 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.[29] The final data set includes about 550,000–650,000 children per birth cohort resulting in about 11 million children overall. More detailed information on the data and the sample is provided in the Appendix section 4.A.1.

4.4.2 Outcome variables

I define measures for physical and mental health using ICD–10 codes. Instead of estimating the effect for about 70,000 diagnoses categorized by the ICD-10 codes, I use broader 2–digit categories. Table 4.1 gives an overview of the considered diagnoses in this study. In addition to the aggregated set of diseases (2-digit level), I provide results for more narrowly defined diagnoses (3- and 4-digit levels of ICD-10 codes). In particular, I select diagnoses captured within the studied sets of diseases that belong to the 50 most frequently reported diagnoses by pediatricians (see ZI, 2015). These are in detail presented in Appendix section 4.A.1.

Similar to van den Berg and Siflinger (2022), I assess three aspects of health: physical health (communicable and non-communicable diseases), mental health, and healthcare consumption. My core physical health measures for communicable diseases capture the following three sets of health conditions: *respiratory diseases* (ICD–10 codes J00–J99), *infections* (ICD-10 codes A00–B99), including any bacterial or viral infection, and *ear problems* (ICD–10 codes H60–H95), capturing diagnoses on the external ear, the middle

[27] 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 obtain an estimate of how many children are covered in the data, I add the official births for the respective birth cohorts that are 0–14 years old for each year between 2009 and 2019.

[28] The number of public health insured children in Germany also includes children who immigrated to Germany, while the official birth records do not include children who immigrated after birth. In contrast, the number on insured children does not include children who emigrated from Germany, while these are included in the birth records. Hence, the estimated share of 86% of children born in Germany might be imprecise as it suffers from the exclusion of emigrating/immigrating children in the official numbers.

[29] As outlined in Appendix section 4.A.1, almost all children appear at least once during the 11-year observation period as they make use of early diagnostic tests.

ear, and the internal ear. These three sets of conditions of communicable diseases 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 be in a causal relationship. In particular, many diagnoses concern contagious diseases common in childhood and often transmit among children; thus, they likely also spread in daycare centers. Many infections may accompany respiratory problems and cause subsequent ear problems. Furthermore, some respiratory diseases or ear problems concerning inflammations could result from infections. Hence, depending on the coding practice of the physician, the three conditions could fall under all three sets of diagnoses. Consequently, to capture all infections, it is essential to study all three groups. As additional measures for physical health (non-communicable diseases), I assess *obesity*, *vision problems*, and *injuries*. Obesity is categorized in ICD-10 codes E65–E68. Additionally, I assess *injuries* (ICD-10 codes S00-S99), which include all kinds of injuries to all body parts (e.g., injuries to the head or knee). Lastly, I consider *vision problems* (ICD-10 codes H00-H59).

To assess the effects on *mental health* including behavioral problems, socio-emotional abilities, and mental health problems, I analyze the effect on ICD–10 codes F00-F99. Note, I measure socio-emotional skills by diagnosed mental and behavioral disorders, which is certainly more extreme than typical measures obtained from survey data (e.g., SDQ-Index). Thus, I rather focus on below-average socio-emotional developments than capturing children's full range of socio-emotional abilities. However, measures in survey data might also underreport socio-emotional development problems as survey respondents (mostly parents) are not professionally trained to recognize behavioral disorders and might have difficulties accepting that their child exhibits behavioral disorders.

To measure communicable diseases and injuries, I compute the annual number of diagnoses per child as a measure of the intensive margin in my main specification (count variables). For chronic conditions (obesity, mental disorders, vision problems), I use the extensive margin, which I construct as binary indicator variables that marks whether a child had a specific diagnosis at least once in a given year. This definition is analogous to Barschkett et al. (2022) who work with the same data.

Lastly, I consider healthcare consumption and healthcare costs. Healthcare consumption is defined as doctor visits 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.[30] Thus, if a child visits two different doctors in a quarter, she has two treatment

[30]Since treatment cases are recorded this way in the data, I cannot define the variable differently for my application.

cases in that specific quarter.[31] I aggregate quarterly cases to the calendar year level, thus counting the number of quarterly treatment cases per year. This means that a patient who visits only the same doctor every quarter would have a yearly count of four treatment cases, irrespective of the actual number of visits to this doctor per quarter. Consequently, the number of treatment cases underestimates the actual number of doctor visits. 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.[32]

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 4.1. 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 about 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 diagnosis at least once per year. In contrast, the likelihood of obesity increases with age, while the prevalence of mental disorders and vision problems increases until age 3–5 and is relatively constant across older age groups. The number of injuries is relatively stable across age groups. 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 treatment cases per year. In line with decreasing treatment cases with age, healthcare costs are higher for younger children (on average 320 Euros per year for 1–2-year-old children) than for older children (249 Euros per year for 9–10-year-old children).

4.4.3 Control variables and daycare coverage rates

The KBV data only includes a few individual-level socio-demographic characteristics, including age, gender, year, birth month, and county of residence. Additionally, county-level ("Landkreise") information, such as the share of migrants and average household income, can be used for heterogeneity analyses. Furthermore, I extract information on the incidence of swine flu at the county level between 2009 and 2011 from the RKI Survstat dashboard (RKI, 2022) as there was considerable regional variation in

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

[32] 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, Kassenärztliche Bundesvereinigung KBV, 2019).

Table 4.1: Outcomes

	1-2 years	3-5 years	6-8 years	9-10 years
Communicable diseases				
Infections (no. per year)	1.39 (1.59)	1.00 (1.29)	0.78 (1.11)	0.66 (1.04)
Infections (prevalence)	0.63 (0.48)	0.53 (0.50)	0.46 (0.50)	0.40 (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)
Non-communicable diseases				
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)
Injury (no. per year)	0.22 (0.56)	0.19 (0.54)	0.19 (0.57)	0.24 (0.65)
Injury (prevalence)	0.17 (0.37)	0.14 (0.35)	0.13 (0.34)	0.16 (0.37)
Vision problems (no. per year)	0.59 (1.16)	0.82 (1.56)	0.79 (1.60)	0.71 (1.45)
Vision problems (prevalence)	0.34 (0.47)	0.38 (0.49)	0.34 (0.47)	0.32 (0.47)
Healthcare consumption				
Treatment cases	6.33 (3.84)	6.14 (4.04)	5.28 (7.46)	4.92 (8.91)
Healthcare costs	320 (313)	287 (320)	245 (393)	249 (450)
Observations	9,042,454	16,840,400	17,167,518	11,674,867

Notes: 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 the share of children who had at least one diagnosis per year. Costs are fee-adjusted.
Source: KBV 2009–2019, own calculations.

the incidence across Germany during the swine flu epidemic.[33] I merge these data to the KBV data and use the swine flu incidences as control variables. The KBV 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. As of 2022, after multiple county reforms that reorganized the counties, there are 401 counties in Germany. Since 2006, the German Statistical Office has provided data on daycare coverage annually.

[33]The swine flu pandemic lasted from 2009 to 2010 (with cases still being prevalent in 2011) and was particularly prevalent among children. For example, in 2009, about half of all swine flu cases occurred in children under 15 (RKI, 2010a). The incidence differed across regions and age groups (e.g., Buda et al., 2010; RKI, 2010b).

Before 2006, only data for 1994, 1998, and 2002 are available. I restrict the analysis to West Germany (323 counties).

4.5 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-differences(DiD) and event study (ES) approaches. Specifically, I compare health outcomes of children born 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). A similar design is also used by, e.g., Havnes and Mogstad (2011b), Müller and Wrohlich (2020), and Bauernschuster et al. (2016). Recent research has identified problems with DiD with staggered implementation utilizing two-way fixed effects and created new estimators to address these issues (e.g., Borusyak et al., 2021; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). I address this issue in section 4.6.3. Below, I describe my main empirical strategy and discuss potential threats to identification and alternative specifications to validate my results.

The daycare expansion generated a steady increase in the coverage rate from the mid-2000s until 2014 on the national level (see Figure 4.1). Specifically, the expansion started in the mid-2000s, peaked between 2008 and 2012, and substantially slowed from 2014 onward. However, Figure 4.2 reveals considerable heterogeneity on the regional (county) level in the total coverage rate and the expansion speed. In my main specification, a generalized DiD, I use the heterogeneous treatment intensity across counties and regress the outcomes directly on the daycare coverage rate in each county controlling for year and county fixed effects and a set of control variables. The advantage of directly regressing the outcomes on the daycare coverage rate is that i) I do not need to make assumptions on the definition of treatment and control group (definition of affected cohorts and counties), and ii) I use the whole variation in treatment intensity across counties. With this approach, I closely follow Müller and Wrohlich (2020).

Specifically, I estimate in a Two Way Fixed Effects (TWFE) framework the following equation:

$$Y_{it} = \alpha_c + \psi_j + \theta cc_{cjt} + X_{it}\beta + \varepsilon_{ijt} \tag{4.1}$$

where Y_{it} represents the health outcomes of child i at age t.[34] cc_{cjt} is the average childcare coverage rate in county c for birth cohort j at age $t \in age[1,2]$[35] with θ being the coefficient of interest. α_c and ψ_j refer to county and birth cohort fixed effects, respectively, and X_{it} is a vector of control variables containing age, gender, and birth cohort dummies interacted with the swine flu incidence in 2009, 2010, and 2011.

Two important aspects must be considered to validly estimate the effect of the day-care expansion within a DiD approach. First, the variance-weighted common trend assumption needs to hold, i.e., in the absence of the expansion, health outcomes across counties should have evolved in parallel conditional on covariates. If the identifying 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 follow Havnes and Mogstad (2011b) and Bauernschuster et al. (2016) and specify a standard DiD framework (Equation (4.A.1)) where the variable of interest is an interaction term between the treatment indicator $Treat_i$ (child lives in a county where daycare expanded a lot) and the reform indicator $Post_i$ (child was born after the expansion). The DiD approach controls for unobserved differences between children from treatment and control counties as well as between children born in different years. In the Appendix section 4.A.2, I outline the DiD framework in more detail. Based on the DiD framework, I provide event study graphs drawing on similar regressions as presented in Equation (4.A.1) in Appendix section 4.A.2 to verify the plausibility of the common trend assumption. In this specification, the $Post$-indicator is replaced by a $Cohort$-indicator, which includes the birth year of children.[36] 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 before 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 4.6.[37] Second, I implement a placebo

[34] In the main specification, I exclude extreme outliers in the dependent variables (count variables), i.e., observations in the top 99.9999 percentile. In robustness checks, I include all observations. The results do not change and are available upon request.

[35] Note, due to data limitations, I only observe child i's county of residence from 2009 onward. Thus, for children born before 2009, I use the county of residence observed in 2009. I use the observed county of residence for all other children when they are 1–2 years old. This assumption is plausible as in my sample, only 8% of children move and only 1% of children move from a treatment to a control county as defined in section 4.A.2.

[36] In an alternative specification, I use the percentage point increase between 2008 and 2012 as a continuous treatment variable. The DiD and event study results are very similar.

[37] Due to data availability, the event-study results are only presented for 6–8-year-old children because, for this age group, a sufficiently large number of pre and post-birth cohorts are available. It seems plausible that the common trend assumption's plausibility can be extrapolated to the other age groups.

regression employing a chronic disease (diabetes mellitus) as an outcome variable, as diabetes should be unaffected by exposure to daycare or other environmental factors.

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 clustered at the county level. This leads to asymptotically valid inference; however, in finite samples like mine (323 clusters), the problem may still be present (Cameron and Trivedi, 2005). Therefore, I additionally estimate wild-bootstrapped clustered standard errors with 9,999 repetitions (Cameron et al., 2008). Under the common trend assumption and the assumption that the marginal effect of an additional daycare slot is constant, θ can be interpreted as the causal effect of an increase in the daycare coverage rate on the outcomes of interest. The parameter θ is an Intention-To-Treat (ITT) parameter, as it measures the effect of increasing access to rather than actual use of daycare.

Because I control for county-specific fixed effects, the daycare expansion does not need to be unrelated to time-invariant county characteristics. However, it is useful to understand the determinants of the expansion across counties. In Appendix Table 4.A.1, I investigate differences in socio-demographic characteristics between treatment and control counties in 2008. The definition of treatment and control group is based on the DiD framework explained in the Appendix section 4.A.2. I can depict only minor differences between treatment and control counties for most characteristics. Interestingly, at the beginning of the expansion period, treatment counties exhibit a slightly higher daycare coverage rate (12.5 vs. 10.7%). Furthermore, the unemployment rate is slightly higher in control counties (5.9 vs. 6.7%), the share of migrants is higher in control counties (6.0 vs. 10.1%), the population density is almost three times as large in control counties compared to treatment counties (299.5 vs. 847.9) and GDP per capita is also higher in control counties (26,435 vs. 33,429 Euros). Thus, in some aspects treatment and control counties are fairly comparable in their socio-demographic composition, while they differ in others. Since the characteristics that exhibit differences are likely stable across the observation period, I control for the differences with the county fixed effects.

4.6 Empirical results

The following section describes and discusses the results of my main specification. Furthermore, I provide evidence for the robustness of my results, present alternative specifications, and show results from heterogeneity analyses. The main results are based on quite broad sets of diseases. Additionally, in Appendix section 4.A.6, I present results for more narrowly defined diagnoses (3- and 4-digit levels of ICD-10 codes) within these sets of diseases.

4.6.1 Generalized DiD results

Table 4.2 reports the results of the daycare expansion on children's health, obtained from estimating Equation (4.1). The first three panels display results for the three sets of communicable diseases: Infections, ear diseases, and respiratory diseases. Panels four to seven show the results for mental disorders, obesity, injury, and vision problems. 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 all children aged 1–10 years. Columns 2 to 5 report age-specific results, i.e., for 1–2 year-old children (column 2), 3–5 year old children (column 3), 6–8-year-olds (column 4), and 9–10-year-olds (column 5). To account for the large number of outcomes and the finite number of clusters, I also report adjusted p-values (known as q-values) for multiple hypotheses testing following Benjamini and Hochberg (1995) to control for the false discovery rate (i.e., the expected proportion of rejections that are type I errors)[38] and wild-bootstrapped clustered standard errors (Cameron et al., 2008).[39] Additionally, the last row in each panel shows the sample mean.

Communicable diseases. For all three sets of communicable diseases, being affected by the expansion leads to a positive instantaneous effect, i.e., an increase in the number of diagnoses for 1–2-year-old children, and negative effects in the long run, i.e., a decrease across older age groups. Specifically, the effects for 1–2-year-old children are highly significant (on the 0.01% significance level) and amount to 0.008 for infections, 0.003 for ear diseases, and 0.016 for respiratory diseases. This means that an increase in the daycare coverage rate by ten percentage points increases the number of diagnoses by 0.08, 0.03, and 0.16, respectively. These estimates correspond to a 5.7% increase for infections, 5.1% for ear diseases, and 5.6% for respiratory diseases compared to the sample means. While the effects on 3–5-year-olds are small and statistically not significant, the results depict negative and significant effects for infections and respiratory diseases for elementary school children (3.9% and 2.2% for 6–8-year-olds and 6.0% and 3.8% 9–10-year-olds, respectively, compared to the sample means). The effects on ear diseases for older children are statistically not significant. The effects when pooling all

[38]Compared to familywise error rate controlling methods such as the Bonferroni correction, this approach has greater power and reduces the penalty to testing additional hypotheses (Anderson, 2008; Benjamini and Hochberg, 1995).

[39]To estimate the wild-bootstrapped clustered standard errors, I apply the R-package *fwildclusterboot*, which is based on the method developed by Roodman et al. (2019). Due to computational power, the estimation only works in the samples for specific age groups.

age groups together (1–10-year-old children) are small and insignificant for infections and respiratory diseases, suggesting that the expansion leads to a shift of illness spells from elementary school age to early daycare age. The total number of infections and respiratory diseases children suffer during their first ten years of life is not affected. For ear diseases, there is a small and positive effect for the pooled age group suggesting that a ten percentage point increase in daycare slots leads to an increase of 3.6% in the number of ear diseases children are affected by between one and ten years.

Non-communicable diseases. The estimates for mental disorders go in the same direction as the estimates for communicable diseases, i.e., an increase in the short run and a decrease in the long run. In particular, there is a positive effect for 1–2-year-old children (5.6%) and negative effects for children in elementary school (3% and 3.6%). For obesity, most point estimates are positive but not statistically significant, with 9–10-year-old children being an exception (decrease of 9.8%). Finally, I obtain mostly insignificant estimates for injuries and vision problems across all age groups.

Healthcare consumption. For treatment cases, the results point out an increase for 1–2-year-olds (1.7%) and a decrease for all other age groups (2% for 3–5-year-olds and 1.1% for 6–8-year-olds). The decreases for older age groups outweigh the increase at age 1–2 years, suggesting that, overall, children between one and ten visit a doctor less often when exposed to daycare (reduction of 1.1%). 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 (except for 3–5-year-olds) and range between -2% (6–8 years) and 4.7% (1–2 years). Overall, there is a slight decrease of 2.8% (1–10 years).

4.6.2 Age-specific results

To better understand the effects of daycare across the age distribution, I plot the coefficients for all age groups separately. Figure 4.3 displays the results age-specific results for infections, ear diseases, respiratory diseases, and treatment cases, i.e., for all outcome variables that prove to be significantly and robustly[40] affected by the expansion. The underlying estimates, as well as the results for the remaining outcomes, are shown in Appendix Table 4.A.2. Similar to the pooled results, estimates for all communicable diseases are positive and significant at ages one to three. From age five or six, estimates turn negative and significant for respiratory diseases and infections, respectively. I only depict significant negative effects for ear diseases at ages four, five, and ten. The effects on treatment cases are positive and significant at age two but

[40]See section 4.6.3 for a discussion on the robustness of the results.

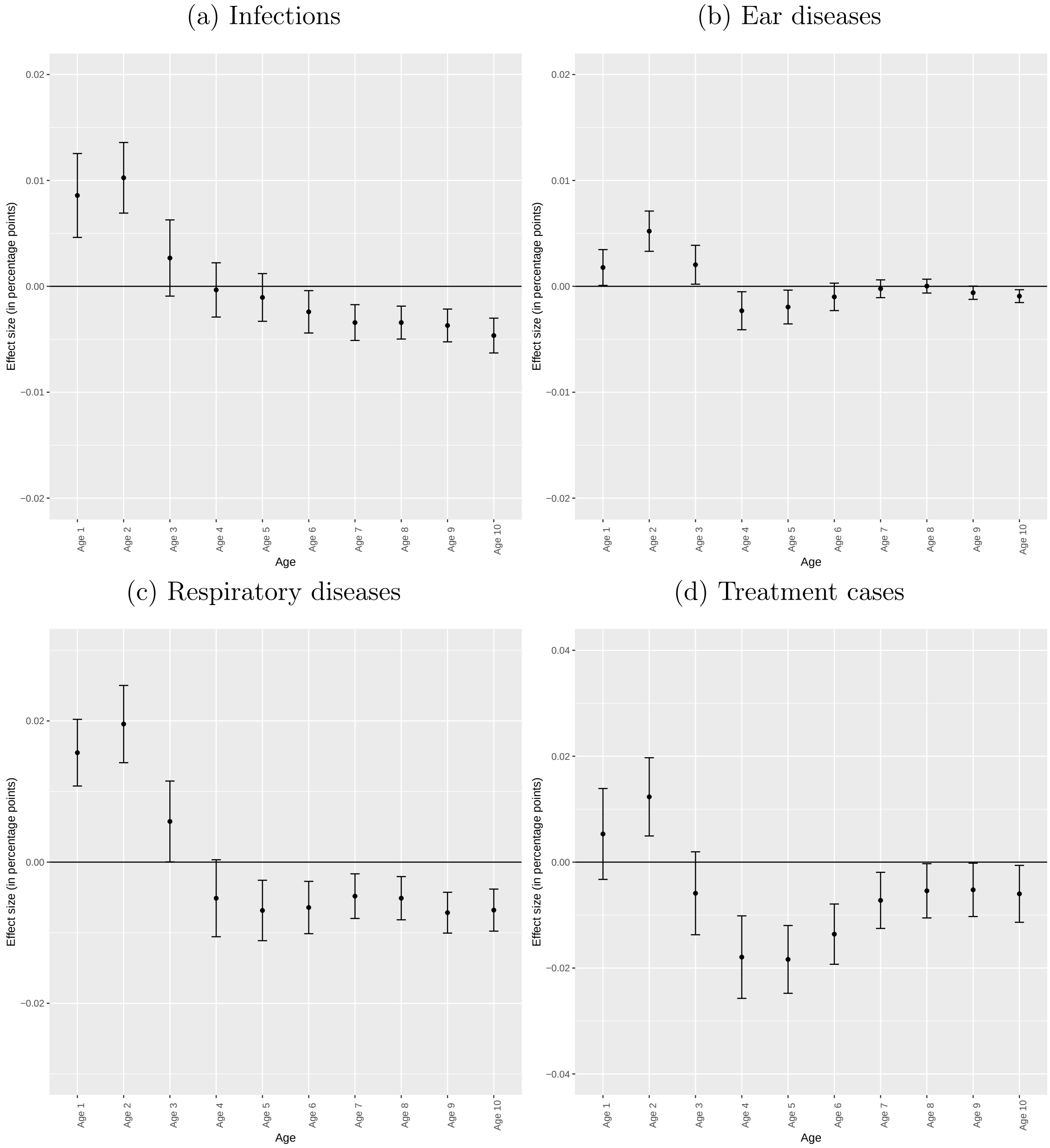
Table 4.2: Generalized DiD Results

	Age: 1-10	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases					
Infections	0.001	0.008***	0.001	−0.003**	−0.004***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
q-value; boot strapped p-value	0.191	0.000; 0.000	0.613; 0.352	0.009; 0.001	0.000; 0.000
Sample Mean (no. per year)	0.924	1.394	1.000	0.777	0.665
Ear diseases	0.002*	0.003***	−0.001	−0.00001	−0.001*
	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0003)
q-value; boot strapped p-value	0.037;	0.001; 0.000	0.619; 0.437	0.977; 0.974	0.045; 0.034
Sample Mean (no. per year)	0.558	0.583	0.84	0.454	0.284
Respiratory diseases	−0.0002	0.016***	−0.001	−0.004*	−0.006***
	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)
q-value; boot strapped p-value	0.910;	0.000; 0.000	0.774; 0.698	0.049; 0.017	0.001; 0.000
Sample Mean (no. per year)	2.207	2.854	2.653	1.852	1.583
Non-communicable diseases					
Mental disorders	−0.001*	0.001**	0.0001	−0.001**	−0.001***
	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0003)
q-value; boot strapped p-value	0.037;	0.011; 0.004	0.774; 0.777	0.022; 0.004	0.000; 0.000
Sample Mean (prevalence)	0.305	0.177	0.37	0.329	0.275
Obesity	0.0002*	0.0001+	0.0001*	0.00005	−0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
q-value; boot strapped p-value	0.037;	0.094; 0.068	0.349; 0.040	0.703; 0.497	0.000; 0.000
Sample Mean (prevalence)	0.031	0.014	0.024	0.033	0.051
Injury	−0.0003*	0.0005*	−0.0003	−0.0001	0.0002
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
q-value; boot strapped p-value	0.060;	0.021; 0.068	0.576; 0.040	0.745; 0.497	0.422; 0.000
Sample Mean (no. per year)	0.204	0.216	0.19	0.189	0.239
Vision problems	0.0002	0.001***	0.0003	−0.0004	−0.001*
	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
q-value; boot strapped p-value	0.426;	1; 0.068	0.613; 0.040	0.216; 0.497	0.017; 0.000
Sample Mean (prevalence)	0.349	0.34	0.381	0.341	0.322
Healthcare consumption					
Treatment cases	−0.006**	0.011**	−0.012***	−0.006*	−0.005+
	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)
boot strapped p-value		0.001	0.001	0.011	0.086
Sample Mean (no. per year)	5.638	6.331	6.135	5.279	4.911
Healthcare costs	−0.754***	1.499***	−0.064	−0.495**	−0.918***
	(0.176)	(0.164)	(0.183)	(0.165)	(0.260)
boot strapped p-value		0.000	0.736	0.001	0.000
Sample Mean (no. per year)	271.106	319.964	287.281	244.466	249.115
Control for age + gender	yes	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes
Birth cohorts	2000-2014	2008-2014	2006-2014	2003-2011	2000-2009
Observations	54,152,621	8,522,322	14,117,165	13,979,553	10,605,774

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in equation 4.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease. q-values are p-values adjusted for multiple hypothesis testing following Benjamini and Hochberg (1995). Boot strapped p-values are calculated based on wild-bootstrapped clustered standard errors accounting for a finite number of clusters.
Source: KBV 2009–2019, own calculations.

negative and significant for all age groups from age four. Interestingly, effects do not fade out with age, but effect sizes are relatively stable for the different age groups.

Figure 4.3: Age-specific results



Notes: The graphs show age-specific effects for 1–10 year old children for infections (panel (a)), ear diseases (panel (b)), respiratory diseases (panel (c)) and treatment cases (panel (d)). The estimates are based on Equation (4.1).
Source: KBV 2009–2019, own calculations.

4.6.3 Robustness

Standard DiD results. Appendix Table 4.A.3 reports the standard DiD results obtained from estimating Equation (4.A.1) for children aged 3–10 years.[41] Generally, the DiD results show a similar picture as the generalized DiD results, i.e., decreases in communicable diseases and treatment cases in the long run and mostly insignificant or small estimates for injuries and vision problems. However, the table shows a different picture for mental disorders, obesity, and healthcare costs compared to the generalized DiD results. For obesity and mental disorders, all coefficients are positive but mostly statistically insignificant. 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 most age groups. Only the estimate for 9–10-year-old children is statistically significant, suggesting an increase in healthcare costs. Since the DiD estimates on obesity, mental disorders, and healthcare costs contradict my main results, I refrain from interpreting the results further.

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 the Appendix in section 4.A.5, which are estimated based on Equation (4.A.2). Figure 4.A.3 presents event-study results for 6–8-year-old children for the three sets of communicable diseases. For infections and respiratory diseases, the point estimates to the left side of the cutoff are mostly insignificant, suggesting a common trend in the absence of expansion. To the right side of the cutoff, the point estimates are negative for both diseases. For respiratory diseases, almost all estimates are statistically significant, while most are only marginally significant or insignificant for infections. This might be explained by a lack of statistical power when estimating the results separately for all birth cohorts. As expected from the insignificant results in the DiD, there is no trend visible for ear diseases – neither before nor after the cutoff. For the other diseases, there is also no trend visible (Figure 4.A.4), suggesting that the common trend assumption is plausible. For treatment cases and healthcare costs, there is also a common trend visible before the reform and in line with the DiD results negative (but insignificant) point estimates after the cutoff for treatment cases (Figure 4.A.5).

To further prove the validity of this assumption, I relax the assumption of constant marginal effects by adding a quadratic term in childcare coverage to Equation (4.1). The estimates point in the same direction as the main results, but standard errors are larger for some estimates, reducing statistical significance (Appendix Table 4.A.4).

[41]Due to data limitations, it is not possible to apply this approach for 1–2-year-old children.

Furthermore, I exclude the $phase-in$ dummy in the regressions for 6–10-year-old children in the DiD. Results do not change compared to the main results (Appendix Table 4.A.5). Finally, I conduct a placebo analysis to provide evidence that my results reflect a reform effect and not just some underlying time trend. I choose diabetes mellitus as an outcome variable, 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 (Appendix Tables 4.A.6 and 4.A.7). To ensure 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 (Appendix Tables 4.A.8 and 4.A.9). In summary, these tests support the plausibility of the common trend assumption, thereby supporting the causal interpretation of my results.

Heterogeneous treatment effects. Recent developments in the DiD literature indicate that TWFE estimators may be subject to biases in staggered treatment implementations under heterogeneity in groups and time. This stems from the fact that the TWFE estimator is a weighted sum of the average treatment effects (ATE) in each group and period. Weights sum to one but individual weights may be both positive and negative (e.g., Borusyak et al., 2021; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). In the context of the daycare expansion, heterogeneity of treatment effects across time is possible, and the staggered rollout might assign negative weights to some treatment effects.[42] Despite the literature's focus on discrete treatment variables, Callaway et al. (2021) point out that the issue of negative weights can also arise in specifications with continuous treatment definitions. Thus, the estimates of my main specification (generalized Did) may be subject to biases arising from negative weights. However, no adjusted estimator is currently available for continuous treatment definitions. In contrast, in my standard DiD framework, by construction, the treatment implementation is not staggered, i.e., the treatment is only implemented at one point in time. Thus, I do not face the problem of time-specific treatment effects. To assess whether the treatment effect varies over time, I estimate the standard DiD specification in Equation (4.A.1) with different definitions of the expansion period. The results displayed in Appendix Tables 4.A.10-4.A.18 are very similar to the baseline results suggesting that varying treatment effects over time are not relevant in this context. Despite being unable to account for negative weights in

[42] In contrast, treatment effect heterogeneity across groups is not relevant as there is an excess demand for daycare slots across the country.

the generalized DiD, I am confident that the potential bias is negligible, as the standard DiD framework largely confirms the generalized DiD results.

Alternative treatment definitions. In my baseline specification of the DiD, I define treatment and control counties by ordering counties by their percentage point increase during the main expansion period and separating the sample at the 30th/70th percentile. In order to test the sensitivity of my results toward this assumption, I run the analysis with other definitions. Specifically, I compare the upper 50% with the lower 50%, the upper 40% with the lower 40%, the upper 35% with the lower 35%, the upper 25% with the lower 25% and the upper 20% with the lower 20%. Finally, I estimate a specification where I use the percentage change increase within the main expansion period instead of a dummy, indicating that a country was above or below a cutoff. Appendix Tables 4.A.19 – 4.A.27 display the results for the different outcomes. The coefficients for ear diseases, injury, and vision problems remain insignificant across most specifications. The results on infections, respiratory diseases, obesity, and treatment cases are, in terms of significance, hardly different from the main results across the different treatment definitions. However, the point estimates increase in magnitude with increasing percentile, i.e., the point estimates are the smallest in the most conservative definition (median separation). This can be explained by the fact that the median separation includes comparing children from counties with very similar expansion rates (just below the median vs. just above the median). In contrast, separation at other percentiles entails the comparison of children from counties that exhibit larger differences in the expansion speed. In line with the uncertain results on mental disorders and healthcare costs from the DiD and generalized DiD estimations, the results are only significant in some specifications.

Extensive/Intensive margin. My main specification investigates the reform's effect on the intensive margin (number of diagnoses) for communicable diseases and injuries as well as the extensive margin for chronic conditions. Alternatively, I investigate the extensive margin for communicable diseases and injuries alongside the intensive margin for chronic conditions. The results are presented in Appendix Tables 4.A.28 and 4.A.29. The direction and significance of the effects are very similar to the main results. Thus, the intensive and extensive margins are affected when considering communicable diseases.[43] This finding suggests that the number of diagnoses and the share of children

[43]Note, the definition of the intensive margin also includes children with zero diagnoses. Thus, in theory, the effects could be driven entirely by the extensive margin (children switching from zero to at least one diagnosis). However, the relative changes (percentage change compared to the sample means) are smaller in the extensive margin than in the intensive margin, suggesting that the main driver of the change in communicable diseases are children experiencing more diagnoses rather than children switching from zero to non-zero diagnoses.

affected by at least one communicable disease per year changes (positive effect at age 1–2 and negative effect at elementary school age). Interestingly, while the extensive margin of vision problems is unaffected, the intensive margin displays significant effects in both the generalized DiD and the standard DiD. Children aged 1–2 years suffer from significantly more vision problems, while children in elementary school are significantly less affected when exposed to daycare. The effects in the intensive margin might be driven by the effect on conjunctivitis as described in Appendix section 4.A.6, which is not a chronic condition but an infectious disease. Thus, counting the number of diagnoses might be a more suitable measure.

4.6.4 Effect heterogeneity

The health effects of the daycare expansion may be heterogeneous across different groups of children. I address this question by exploring whether the effects of the expansion on health outcomes are heterogeneous by gender and by areas with different socio-economic status.

Gender. Appendix Table 4.A.33 shows the results of estimating Equation (4.1) for girls and boys separately. For most outcomes, there are no gender differences in the impact of the expansion on health. The exception presents injuries: While girls have, in general, a lower prevalence of injuries than boys (e.g., 0.172 injuries per year at age 6–8 years for girls vs. 0.206 injuries per year for boys), entering daycare earlier significantly increases the risk of injuries for girls at age 1–2 years while there is no such effect for boys.[44] The effect corresponds to a 5.1% increase following a ten percentage point increase in the coverage rate.

Socio-economic background. Daycare in Germany is universal and open to everyone. However, the take-up rate of the scarce daycare slots for children below the age of three is higher among highly educated and non-migrant families (e.g., Jessen et al., 2020). Many large-scale early childhood interventions are found to benefit more disadvantaged populations (see Almond et al. (2018) for a review) and I now turn to study whether the expansion's impacts vary by socio-economic background. Due to data limitations, I do not observe socio-economic characteristics at the individual level, but I can observe average household income and the share of migrants at the county level. To construct a measure of the income level (share of migrants), I sort the counties by their average

[44]To test whether the effects are statistically significantly different between the groups, I add an interaction term between *cc* and a binary variable indicating whether the child is male in a separate regression based on equation 4.1. The coefficient on this interaction term is significant for injuries at age 1–2 years.

household income (share of migrants) and separate the sample at the top/bottom 30th percentile to compare counties with high and low household income (share of migrants).

The results in Appendix Tables 4.A.34 and 4.A.35 reveal that children from lower socio-economic background (low-income counties, counties with high shares of migrants) drive the results on communicable diseases. Specifically, the effects on infections are more pronounced for children from low-income counties while the effects on ear and respiratory diseases are larger for children from counties with high shares of migrants. Similarly, early daycare reduces obesity in elementary school children from counties with high shares of migrants and leads to an earlier detection of vision problems in young children from low-income counties.[45] There are no sizable differences between high and low-income counties and between counties with high and low shares of migrants for the remaining outcomes. The absence of differences by socio-economic status in some outcomes (e.g., mental disorders) could be explained by the fact that even in low-income areas and areas with high shares of migrants, more highly educated and non-migrant families take advantage of the supply of daycare slots. Another explanation could be differences in daycare quality (e.g., differences in group sizes) between more and less advantaged counties. Individual-level information on the socio-economic background could provide more precise estimates of the socio-economic differences of the impact of the reform.

4.6.5 Discussion of the results

In sum, the results on communicable diseases provide evidence of a daycare-driven intertemporal substitution of illness spells from the first years of the elementary school towards the first years of daycare. More precisely, on the one hand, children suffer more frequently from infections, ear, and respiratory diseases when entering daycare at age 1–2. On the other hand, at elementary school age, children fall sick less often with these conditions. Interestingly, in total, between ages one and ten, children who enter daycare earlier suffer from the same number of infections and respiratory diseases as children who enter daycare later. In contrast, there is a small positive effect for the overall age group for ear diseases, suggesting that children who enter daycare at age 1 or 2 suffer from more ear diseases up until age ten than children who enter daycare later. The results are intuitive as children in daycare are in close contact with other children and, therefore, exposed to many viruses and bacteria. Exposure to viruses and

[45]The observed differences are statistically significant for infections, ear diseases and obesity but not for respiratory diseases and vision problems.

bacteria leads to worse health in the short run but initiates the immunization process earlier, leading to fewer infections in the longer run.

My results are in line with the hygiene hypothesis, as well as the results from van den Berg and Siflinger (2022), Cattan et al. (2021) and the medical literature (e.g., Enserink et al., 2013).[46] The results of van den Berg and Siflinger (2022) are similar in that they also find increases in infections and respiratory diseases following daycare exposure in the short run and better health in the long run. Their study finds more pronounced effects on ear diseases in the long run, while my effects mainly hold for infections and respiratory diseases. However, the three sets of conditions are closely related and different reporting practices and daycare environments in Sweden and Germany could explain the differences. Furthermore, my analysis of the more narrowly defined outcomes also reveals effects of the expansion on a subgroup of ear diseases, namely otitis media. Similarly, Cattan et al. (2021) provide evidence that exposure to Sure Start, among other things also entailing daycare, leads to an increase in hospital admissions due to infectious illnesses at age one and a decrease in later childhood and adolescence. In contrast to my results, there is evidence that a large-scale daycare reform in Quebec led to adverse effects on health both in the short and long run. The effects are mainly driven by children who had access to daycare at very young ages (Baker et al., 2008, 2019; Kottelenberg and Lehrer, 2013). Differences in daycare quality might drive these contrasting results: While daycare in Sweden is considered high-quality (e.g., Bremberg, 2009), the expansion in Quebec was relatively cheap and is considered low-quality care (e.g., Kottelenberg and Lehrer, 2013). Findings from other studies analyzing smaller and targeted programs that entail daycare but also other components, such as home visits (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 better understand the effect sizes, I compare my estimates with estimates found for other factors that influence health. Age, for example, is a critical determinant of infections, ear, and respiratory diseases (compare Table 4.2). Additionally, education, air pollution and second-hand smoke are known to be critical factors influencing child and adolescent health (e.g., Coneus and Spiess, 2012b; Hawkins et al., 2016; Huebener, 2022). Comparing effect sizes from the literature with my effects reveals that the observed increase in 1–2-year-old children and reductions in elementary school-aged children in the prevalence of communicable diseases appear sizable. One additional year

[46]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.

of education leads to three to four times bigger effect sizes than a ten percentage point higher daycare coverage rate at age 1–2 years (e.g., Huebener, 2022). In comparison, one year in age or smoke-free legislation leads to about twice as large effect sizes for communicable diseases (e.g., Table 4.2, Hawkins et al., 2016). Specifically, Hawkins et al. (2016) denote reductions of 8–12 percent in emergency department visits of children associated with asthma, respiratory infections, and ear infections following smoke-free legislation.

My results provide little evidence that the daycare expansion affects mental health. In Appendix section 4.A.6, 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 by the reform might suffer more often from development disorders at young ages but less often at elementary school age. However, I do not find significant and robust effects for other subgroups. My findings contrast van den Berg and Siflinger (2022), who point out 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 daycare for all age groups, the expansion in Germany caused a switch from mainly home care into daycare only for the children below the age of three. Children in informal daycare could benefit from a switch to formal daycare where care actors are potentially more qualified. In Germany, almost all children from age three onward were in daycare before the reform. Therefore, only the timing of entering daycare changed, not the daycare environment.

My analysis suggests that entering daycare does not affect the prevalence of obesity, vision problems, or injuries. The null effects on vision problems contrast findings from Hong et al. (2019), who suggest that attending a pre-kindergarten program in the US increases the probability of being diagnosed with vision problems. One reason for the different findings could be that the findings in Hong et al. (2019) relate to children from low-income families. At the same time, the expansion in Germany, in principle, affected all children, but particularly children from higher socio-economic backgrounds (Jessen et al., 2020). My heterogeneity analysis supports this argument, as in low-income areas, children are more likely to be diagnosed with vision problems at age 1–2 years.

According to my results, the daycare expansion has no clear effects on the prevalence of obesity. However, I observe a reduction in obesity among elementary school aged children from counties with a high share of migrants. Lauber (2015) points out that children at the margin, i.e., children whose daycare usage is affected by regional daycare provision, gain from enrollment at 30 months or earlier. Specifically, they show 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. Lauber (2015) and D'Onise et al. (2010) study daycare attendance of pre-school aged children. In their cases, children who do not attend daycare may not attend daycare at all before entering school. In my study, the expansion affects daycare attendance of 1–2-year-old children and almost all children attend daycare from age three. These age differences might explain the differences in the results.

In line with more communicable diseases at young ages and fewer at older ages, I provide evidence that the daycare expansion led to more healthcare consumption at ages 1–2 and less healthcare consumption in the long run. This finding matches the results of van den Berg and Siflinger (2022). Similarly, Cattan et al. (2021) find more hospital admissions in the short run but fewer in later childhood and adolescence. However, the effects on healthcare consumption are relatively small in magnitude (+1.7% for 1–2 years and −2% and −1.1% for 3–5 years and 6–8-year-olds, respectively for the generalized DiD results assuming a ten percentage point increase in the daycare coverage rate).[47] 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 are 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 detect sizable effects of the expansion on healthcare costs. This result is quite surprising, but a potential explanation lies in the billing system of the 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.

The results from my heterogeneity analysis, namely the absence of gender difference for most outcomes and more pronounced effects for children from disadvantaged areas, align with the findings from the literature (Almond et al., 2018; Bosque-Mercader, 2022; Hong et al., 2019; van den Berg and Siflinger, 2022). However, in terms of gender differences, some studies find daycare to be more beneficial for boys (e.g., the literature on targeted early childhood interventions and evidence on the Sure Start program in the UK, Carneiro and Ginja, 2016; Cattan et al., 2021; Conti et al., 2016; Gray-Lobe et al., 2021).

[47]For comparison, the increase in respiratory diseases at age 1–2 corresponds to 5.6%.

4.6.6 Implications of the results

My results raise the question of whether the substitution of illness spells of infections and respiratory diseases from the first years of elementary school to the first years of daycare is beneficial from a welfare perspective. 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 of the reform in terms of health, other aspects such as duration of illness spells at different ages, sickness absence at school/daycare, severity and long-term health effects, and spill-over effects to siblings or parents need to be considered.

Severity and long-term health effects. On the one 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. In turn, higher antibiotic intake in children may also have adverse long-term effects on cognitive development and other health outcomes such as obesity (e.g., Baron et al., 2020; Mbakwa et al., 2016). Similarly, the medical literature mostly finds adverse long-term effects (e.g., an increase in asthma) following severe respiratory infections in children below 12 months (e.g., Carraro et al., 2014). However, as daycare in Germany starts for most children earliest when they turn one year old, this potential channel is less relevant in this study.

On the other hand, the effect sizes for infections and respiratory diseases at elementary school age are relatively stable and negative across older age groups suggesting that there are effects beyond age ten. A small strand of the literature evaluates long-term health effects of universal daycare reforms and find mixed results (e.g., Baker et al., 2019; Bosque-Mercader, 2022; Breivik, 2020; Haeck et al., 2018). For example, Bosque-Mercader (2022) finds a lower prevalence of asthma following a daycare expansion in Spain in young adults aged 11–27 years. Similarly, Haeck et al. (2018) find that the increase in asthma prevalence following the Quebec expansion is offset in the long-run. My results (Appendix section 4.A.6) also suggest an increase in asthma in the short-run and a decrease in the long-run. The evidence on long-run health is mixed concerning other health outcomes, such as healthcare consumption. Taken together, the stable negative coefficients on communicable diseases across elementary school age combined with the literature finding improved long-term health outcomes suggest that the impact of the reform is not limited to the time horizon studied in this paper but may reach adolescence and adulthood. These potential improvements in health beyond the study period highlight the benefits of early daycare attendance.

Spill-over effects to siblings and parents. In 2010 in Germany, the average age difference between the first and second child was about four years (Pötzsch, 2012). Thus, if children enter daycare before the age of three, it is more likely that they do not have younger siblings yet. As shown by Daysal et al. (2022), younger siblings have a significantly higher likelihood of being hospitalized before age one for respiratory conditions and to experience worse long-run outcomes in terms of health, education and labor market success than older siblings. They argue that one explanatory channel is older siblings "bringing home" infections from daycare. Therefore, if older siblings catch infections below three years when they do not yet have siblings, this potentially improves the health outcomes of younger siblings. Unfortunately, as I cannot link siblings in my data, this empirical question remains to be answered by future research.

To investigate potential spillover effects on parents, I conduct an additional analysis drawing on data from the German Socio-Economic Panel (SOEP).[48] I construct a sample including children between one and ten years observed between 2009 and 2019, including the socio-economic characteristics of the parents. As outcome variables, I use mothers' and fathers' general health, the number of doctor visits in the past three months, the number of days missed at work due to sickness in the previous year, and the number of days missed at work due to sickness of the child in the previous year. In a simple OLS framework, I regress the outcome variables on the explanatory variable, a variable indicating if a child attended daycare when she was below three years, and a set of control variables.[49] The results are presented in Table 4.A.36 and can be interpreted as associations between children's early daycare attendance and parental health. However, due to endogeneity issues, the results of this additional analysis do not provide causal estimates.

Specifically, there is evidence of a negative association between parental health and children's daycare attendance when children are 1–2 years old (Table 4.A.36). This suggests that also parents suffer from infections young children "bring home" from daycare. In line with child health improving with age, fathers of children of age 6–8 years also seem to benefit from better health when their children enter daycare early. In contrast, there is no such correlation visible for mothers. However, for mothers of young children, sickness absence at work due to illness (of their child) is positively correlated with children having been in daycare when they were below three years. Similarly, there is a negative relationship between the number of days missed at work due to the child's illness when these children are 3–5 years old. No such effects are

[48]More information on the SOEP can be found in Goebel et al. (2019).

[49]The set of control variables includes parental education, survey year, cohabitation status, birth order, parental labor force status, parental migration background, household income, parental age, child sex, the federal state of residence, age of siblings and all-day daycare/school attendance.

visible for fathers, which is in line with mothers bearing most of the care work for young children in Germany. Generally, the results provide evidence that parents' illness load and sickness absence at work also increase in the short-run and decrease in the long run. As maternal hours worked increase with child age (Federal Institute for Population Research, 2020), decreasing sickness absence when children are older may increase productivity, thereby enhancing welfare.

Sickness absence at school/daycare. Shifting illness spells may also entail a shift in the timing of absenteeism at school or daycare. The reduction in infectious diseases at elementary school age combined with mothers' reduced sickness absence at work due to child health when children are at elementary school age suggests that children's sickness absence at elementary school decreases. On the one hand, substituting infections from elementary school to daycare age might be desirable as the sickness absence of students in schools may be reduced. Sickness absence at school is associated with worse educational and labor market outcomes (e.g., Cattan et al., 2017), emphasizing the benefit of reducing sickness absence at school. On the other hand, sickness absence at daycare centers could disrupt the relationship between children and daycare teachers as well as fellow children in daycare, thus harming children's early development. Future research could investigate the trade-off between sickness absence in school and daycare to find the "optimal" timing for infections.

Duration of illness spells. Differences in the duration of illness spells by age may provide arguments in favor of or against the change in the timing of illness spells. I do not observe illness/symptom duration in my data. Instead, Thompson et al. (2013) provide an overview of the duration of infectious illnesses (e.g., common cold symptoms, respiratory tract infection symptoms, earache, sore throat, cough) obtained from various medical studies. Age patterns vary by symptoms. For example, common cold symptoms resolve on average after 1.5 weeks in infants, 2.1 weeks in three-year-old children, and 1.3 weeks in seven-year-olds. For respiratory tract infections, time to symptom resolution is relatively similar from infants to elementary school-aged children (about 6-9 days, depending on the study). Thus, the medical literature does not provide strong evidence that illness duration varies substantially by age. Whether there is also no age pattern in this particular setting remains an empirical question for future research.

4.7 Conclusion

This paper provides novel insights into the causal effects of a large-scale daycare expansion for children below three 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 approaches.

My empirical results, based on administrative health records from Germany, provide evidence that early daycare attendance does not affect the illness load of children overall, but leads to a substitution of illness spells (respiratory, ear and infectious diseases) from elementary school age to the first years of daycare. The observed change in communicable diseases appears sizable in light of other reforms or factors that affect health. For example, one year in age or smoke-free legislation produces about twice as large effect sizes. I do not find significant and robust changes in mental health or obesity. For injury and vision problems, I detect null effects. Healthcare consumption increases at ages 1–2 years while it decreases at ages 3–5 and 6–8 years. Despite changes in the prevalence of diagnoses and the number of doctor visits, there is no clear evidence of the effects on healthcare costs. The findings are robust to a large set of robustness checks such as different definitions of the treatment status and the expansion period, the application of multiple hypothesis testing methods to obtain valid p-values, and plausibility checks of the common trend assumptions.

Heterogeneity analysis indicates no gender differences in the expansion's impact but more pronounced effects for children from disadvantaged areas. This finding aligns with consistent evidence that children from disadvantaged backgrounds can benefit disproportionately from access to daycare in a range of dimensions, including health (e.g., Almond et al., 2018). The impact of the expansion likely works through several mechanisms, including the earlier onset of an immunization process (hygiene hypothesis), formation of health habits, formation of socio-emotional and cognitive skills, and changes in the child's environment other than daycare attendance *per se* (e.g., increased maternal labor market participation).

Evidence from additional analysis and the literature reveals that the effects are not bound to children directly affected by the expansion but may spill over to siblings and parents. Namely, an additional analysis using survey data (SOEP) indicates that parents of children who enter daycare before the age of three suffer from worse health in the short run but benefit from improved health when children are older. Similarly, sickness absence from work decreases with the age of the children, which may increase productivity (maternal labor market participation increases with child age), thereby enhancing welfare. Additionally, younger siblings may benefit from older siblings entering daycare earlier as the shift of illness spells may reduce the number of infections older siblings have and "bring home" after younger siblings are born (Daysal et al., 2022). Furthermore, other factors such as long-term health effects, duration of illness spells at different ages or sickness absence at school or daycare do not provide evidence that changing the timing of infections to earlier years leads to detrimental effects that would challenge the daycare entry age of children. Assessing and contrasting the costs and (health) benefits of the reform, including a more precise analysis of spill-over effects on parents and siblings, are avenues for future research. Additionally, relying on additional data sources such as prescription and inpatient registers could shed light on the effects of early daycare attendance on other health dimensions including severe illness treated in hospitals and antibiotic intake.

4.A Appendix

4.A.1 Additional information on the data

Procedure to balance the sample

Analogous to Barschkett et al. (2022), first I create variables indicating the number of times an outcome, for example, respiratory conditions, was diagnosed in a specific period. Additionally, I create variables that measure the extensive margin, i.e., indicate if a patient had at least one relevant diagnosis per year. Secondly, I aggregate the data to a yearly level so 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.

Detailed information about the number of children in the data

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) states 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 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. In the dataset, children are identified via a unique patient ID 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 systematic, thusly not threatening my identification strategy.

Detailed information on outcome variables

Respiratory diseases. 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), Acute laryngitis, tracheitis (J04) and Chronic rhinitis, nasopharyngitis and pharyngitis (J31) (ZI, 2015).

Infectious diseases. Common infections occurring in childhood are Other infectious diseases (B99), Viral infection of unspecified sites (B34), and Other gastroenteritis and colitis of infectious and unspecified origin (A09) (ZI, 2015).

Ear problems. Here, the most common diagnoses occurring among children are Suppurative and unspecified otitis media (H66), Other hearing loss (H91), and Non-suppurative otitis media (H65) (ZI, 2015).

Vision problems. I study Conjunctivitis (H10), Visual disturbances (H53), and Visual impairment, including blindness (binocular or monocular) (H54), which belong to the most frequent vision diagnoses among children in Germany (ZI, 2015).

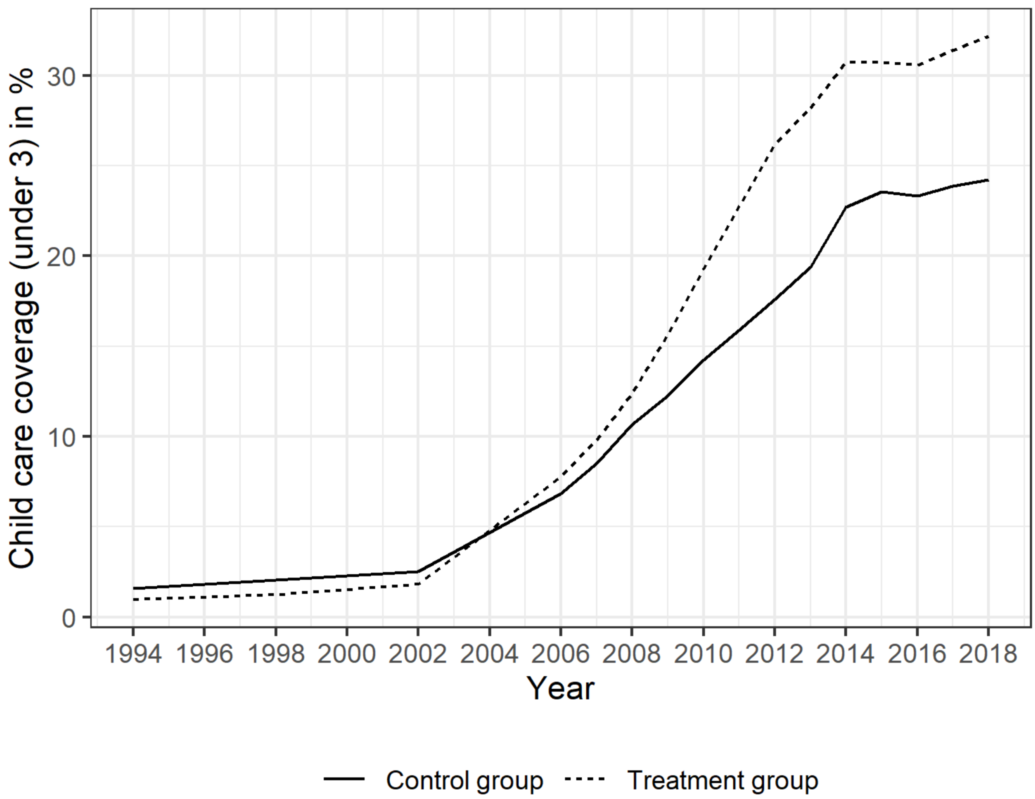
Mental health. The most frequent mental health diagnoses among children in Germany are Specific developmental disorders of speech and language (F80), Hyperkinetic disorders (attention deficit/hyperactivity disorder (ADHD), F90), Unspecified disorders of psychological development (F89), Specific developmental disorders of motor function (F82), Other behavioral and emotional disorders with onset usually occurring in childhood and adolescence (F98), and Mixed specific developmental disorders (F83) (ZI, 2015).

4.A.2 Standard difference-in-differences framework

In order to be able to test for the plausibility of the common trend assumption, I follow Havnes and Mogstad (2011b) and Bauernschuster et al. (2016) and specify a standard DiD framework. I define 2008–2012 as the main expansion period. Defining 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 greatest 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. The expansion did not affect cohorts born before 2005 (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 groups, I order counties according to the percentage point increase in daycare coverage rates from 2008–2012. This definition allows for a clear treatment definition that does not require applying DiD estimators (e.g., Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020) for staggered treatments. I then separate the sample at the 30th/70th percentile, the upper 30% constituting the treatment counties and the bottom 30% the control group. Children from counties between the 30th and 70th percentile are excluded from the analysis. Figure 4.A.1 depicts daycare coverage rates before, during and after the expansion in treatment and control counties. The graphs move almost in parallel until 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 cutoffs other than the 30th/70th percentile (e.g., median).

Figure 4.A.1: Daycare expansion during 1994 and 2018



Notes: 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 regression model, estimated by OLS, can be defined as

$$\begin{aligned} Y_{it} = \psi_t + \gamma_1 Treat_i + \gamma_2 Post_t + \gamma_3 Phasein_t + \gamma_4 (Treat_i \times Phasein_t) \\ + \theta(Treat_i \times Post_t) + X_{it}\beta + \varepsilon_{it} \end{aligned} \tag{4.A.1}$$

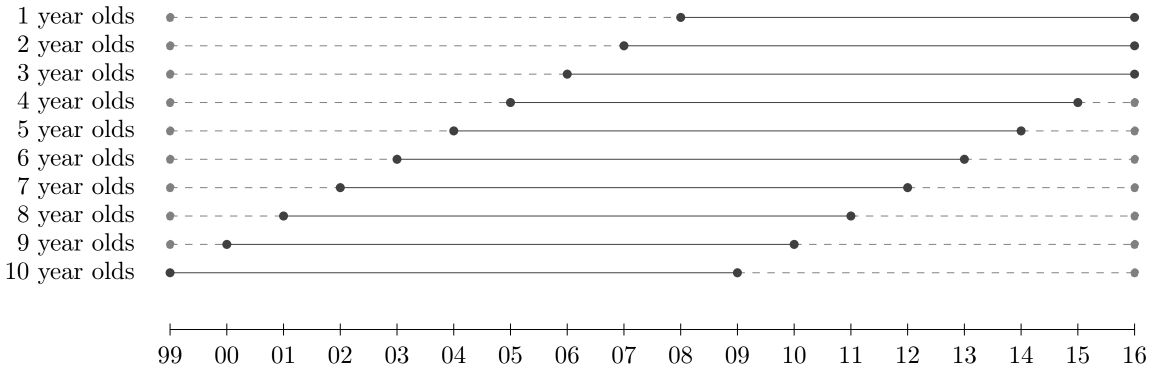
where $Treat_i$ is a dummy variable that indicates whether child i lived in a treatment county. In an alternative specification, I use the percentage point change between

2008 and 2012 as a continuous treatment variable. This continuous treatment variable reduces the information loss in the standard DiD design and relaxes the assumption of the treatment status of counties. $Phasein_t$ is 1 if child i was born in year $t \in [2005, 2006]$[50] and $Post_t$ turns 1 if child i was born in year $t \in [2007, 2011]$. All other variables are the same as in Equation (4.1). Interacting $Treat_i$ and $Post_t$ marks all children affected by the expansion, i.e., children born between 2007 and 2011 and living in a treatment county. Thus, θ is the coefficient of interest and captures the expansion's intention-to-treat (ITT) effect. I interpret this as an ITT effect, as my model estimates the reduced form impact on all children from post-reform cohorts who reside in the treatment area. The benefit of estimating an ITT effect is that it captures the full reform impact. Thus, not only is the effect on treated children 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, Barschkett et al., 2022a). However, the ITT averages the effect over all children in treated counties. Therefore, the effect size is difficult to interpret and needs to be weighed against the size of the expansion. 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 (4.A.1) with the daycare coverage rate in county c for birth cohort t on the left-hand side. θ then gives 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.

As the KBV data are only available from 2009 onward, the standard DiD approach can only be applied to children three years and older. This is because a pre-period is missing for younger children, as only birth cohorts from 2007/08 onward are observed. Figure 4.A.2 in the appendix shows data availability for the different age groups and birth cohorts. As there is 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.

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

Figure 4.A.2: Data availability by age group and birth cohort



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

Source: KBV 1999–2016, own calculations.

4.A.3 County characteristics

Table 4.A.1: Descriptive statistics treatment vs. control counties

	Control counties (N = 97)	Treatment counties (N = 97)	P-Value
Daycare coverage rate			
mean (sd)	10.654 (5.084)	12.522 (4.871)	0.010
Unemployment rate			
mean (sd)	6.699 (3.041)	5.906 (2.362)	0.044
Share of population U3			
mean (sd)	2.498 (0.190)	2.385 (0.2256)	0.000
Average age			
mean (sd)	42.479 (1.110)	42.745 (1.2860)	0.125
Share of migrants			
mean (sd)	10.081 (4.701)	5.967 (2.635)	0.000
Fertility rate			
mean (sd)	1.397 (0.109)	1.403 (0.105)	0.732
Infant mortality			
mean (sd)	3.622 (1.967)	3.474 (2.099)	0.614
Life expectancy			
mean (sd)	80.061 (1.001)	79.980 (0.787)	0.536
Female employment rate			
mean (sd)	44.294 (3.948)	45.991 (3.397)	0.002
Household income			
mean (sd)	1,604.526 (209.639)	1,575.454 (168.331)	0.288
Population density			
mean (sd)	847.856 (836.735)	299.536 (440.367)	0.000
GDP per capita			
mean (sd)	33.429 (12.006)	26.435 (12.118)	0.000
Excess nitrogen			
mean (sd)	79.434 (24.904)	70.934 (27.067)	0.024

Notes: Means (standard deviations) and p-values testing for the difference between the groups are reported.
Source: INKAR 2008, own calculations.

4.A.4 Results by age group

Table 4.A.2: Generalized DiD Results by age group

	Age: 1	Age: 2	Age: 3	Age: 4	Age: 5	Age: 6	Age: 7	Age: 8	Age: 9	Age: 10
Communicable diseases										
Infections	0.009^{***}	0.010^{***}	0.003	-0.0003	-0.001	-0.002^{*}	-0.003^{***}	-0.003^{***}	-0.004^{***}	-0.005^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ear diseases	0.002^{*}	0.005^{***}	0.002^{*}	-0.002^{*}	-0.002^{*}	-0.001	-0.0002	0.00002	-0.001^{+}	-0.001^{**}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Respiratory diseases	0.016^{***}	0.020^{***}	0.006^{*}	-0.005^{+}	-0.007^{**}	-0.006^{***}	-0.005^{**}	-0.005^{**}	-0.007^{***}	-0.007^{***}
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Non-communicable diseases										
Mental disorders	0.001	0.001^{***}	0.001	-0.0002	-0.001^{+}	-0.001^{**}	-0.001^{*}	-0.001^{**}	-0.001^{***}	-0.001^{***}
	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Obesity	0.0001^{+}	0.0001^{+}	0.0001^{+}	0.0001^{+}	0.0001	0.00002	0.0001	− 0.0001	-0.0003^{**}	-0.001^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Injury	-0.0003	-0.0003	-0.0004^{+}	-0.0001	-0.0003	-0.0003	-0.0002	0.0001	0.0002	0.0004^{+}
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Vision problems	0.001^{***}	0.002^{***}	0.001^{**}	-0.0001	-0.0003	-0.0002	-0.0002	-0.001^{*}	-0.0004^{+}	-0.001^{*}
	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Healthcare consumption										
Treatment cases	0.005	0.012^{**}	-0.006	-0.018^{***}	-0.018^{***}	-0.014^{***}	-0.007^{**}	-0.005^{*}	-0.005^{*}	-0.006^{*}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Healthcare costs	0.002^{*}	0.005^{***}	0.002^{*}	-0.002^{*}	-0.002^{*}	-0.001	-0.0002	0.00002	-0.001^{+}	-0.001^{**}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Control for gender	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Control for swine flu	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Birth cohorts	2008-2018	2007-2017	2006-2016	2005-2015	2004-2014	2003-2013	2002-2012	2001-2011	2000-2010	1999-2009
Observations	4,287,667	4,754,773	5,278,596	5,801,293	5,760,578	5,725,600	5,708,062	5,733,983	5,806,102	5,868,892

Notes: $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Robust standard errors clustered on county-level are in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.
Source: KBV 2009–2019, own calculations.

4.A.5 Robustness

Table 4.A.3: DiD Results

	Age: 3-10	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases				
Infections	-0.004	-0.013	-0.027^{**}	-0.009
	(0.011)	(0.012)	(0.010)	(0.009)
q-value; boot strapped p-value	0.738; 0.734	0.581; 0.265	0.023; 0.004	0.581; 0.313
TT	−0.8%	−3.4%	−6.1%	−2.7%
Ear diseases	0.016^{+}	0.005	-0.006	0.001
	(0.008)	(0.008)	(0.005)	(0.003)
q-value; boot strapped p-value	0.118; 0.041	0.581; 0.487	0.337; 0.183	0.524; 0.683
TT	5.7%	1.3%	−2.3%	0.6%
Respiratory diseases	-0.031	-0.025	-0.050^{*}	-0.024
	(0.021)	(0.020)	(0.020)	(0.016)
q-value; boot strapped p-value	0.257; 0.135	0.581; 0.196	0.036; 0.005	0.684; 0.144
TT	−2.7%	−2.1%	−4.5%	2.8%
Non-communicable diseases				
Mental disorders	0.003	0.007^{*}	0.004	0.002
	(0.004)	(0.003)	(0.003)	(0.003)
q-value; boot strapped p-value	0.478; 0.406	0.128; 0.013	0.370; 0.258	0.581; 0.495
TT	1.7%	4.6%	2.3%	1.5%
Obesity	0.003^{***}	0.001	0.002^{**}	0.002^{*}
	(0.001)	(0.001)	(0.001)	(0.001)
q-value; boot strapped p-value	0.006; 0.000	0.581; 0.350	0.014; 0.001	0.208; 0.025
TT	12.9%	10.3%	11.0%	7.6%
Injury	-0.002	-0.001	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
q-value; boot strapped p-value	0.277; 0.192	0.635; 0.627	0.699; 0.603	0.581; 0.426
TT	−1.5%	−1.3%	−0.9%	−1.5%
Vision problems	-0.006^{*}	-0.002	-0.001	0.002
	(0.003)	(0.003)	(0.003)	(0.003)
q-value; boot strapped p-value	0.118; 0.034	0.581; 0.483	0.802; 0.799	0.581; 0.434
TT	−3.0%	−1.3%	−0.5%	1.1%
Healthcare consumption				
Treatment cases	0.001	-0.062^{**}	-0.062^{*}	0.015
	(0.031)	(0.023)	(0.029)	(0.029)
boot strapped p-value	0.946	0.004	0.028	0.593
TT	0.03%	−2.6%	−1.7%	0.6%
Healthcare costs	5.840^{+}	-0.319	3.016	9.016^{**}
	(3.425)	(1.302)	(2.126)	(2.965)
boot strapped p-value	0.080	0.785	0.147	0.002
TT	3.9%	−0.3%	2.1%	6.8%
Control for age + gender	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2000-2011	2006-2011	2003-2011	2000-2009
First stage	59.7%	38.8%	57.2%	57.2%
Observations	21,215,410	5,235,062	7,903,346	5,990,518

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.A.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. q-values are p-values adjusted for multiple hypothesis testing following Benjamini and Hochberg (1995). Boot strapped p-values are calculated based on wild-bootstrapped clustered standard errors accounting for a finite number of clusters. The ITT is calculated by scaling the coefficient with the pre-treatment mean. The TT is calculated by diving ITT/first stage, where the first stage estimates amount to 59.7% for the age group 3–10, 38.8% for 3–5, 57.2% for 6–8 and 52.7% for 9–10. The coefficients show the effect of living in a fast-expanding county and being born after the reform on the respective disease.
Source: KBV 2009–2019, own calculations.

Event-Study graphs

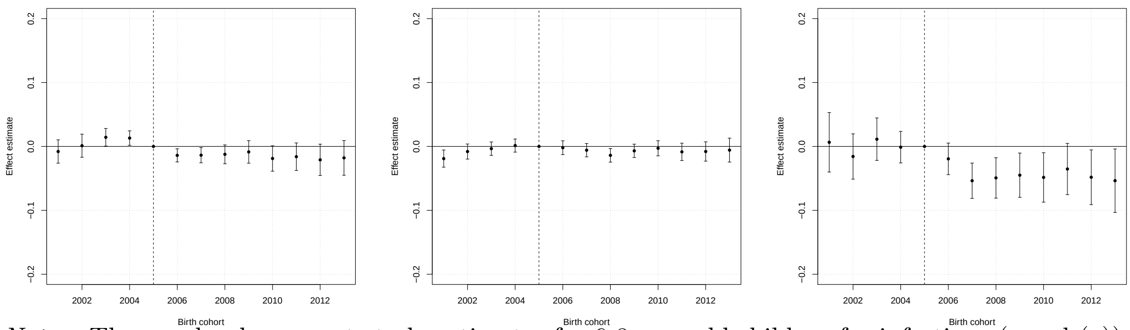
The event-study graphs are estimated based on the following specification

$$Y_{it} = \psi_t + \theta(Treat_i \times Cohort_i) + X_{it}\beta + \varepsilon_{it} \tag{4.A.2}$$

where $Cohort_i$ represents the birth year of child i, where 2005 serves as the reference cohort, all other variables are the same as in Equation (4.A.1).

Figure 4.A.3: Event study: Communicable diseases

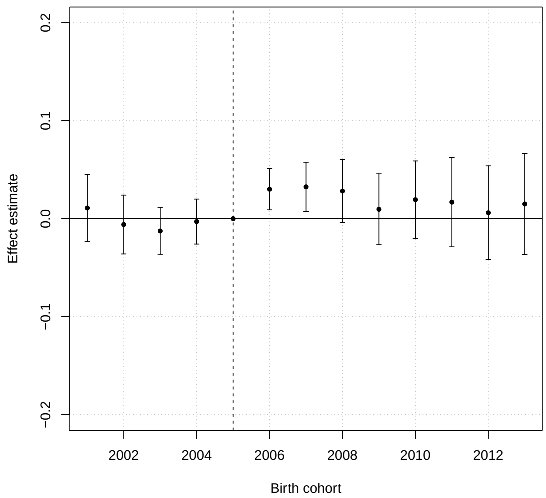
(a) Infections (b) Ear diseases (c) Respiratory diseases



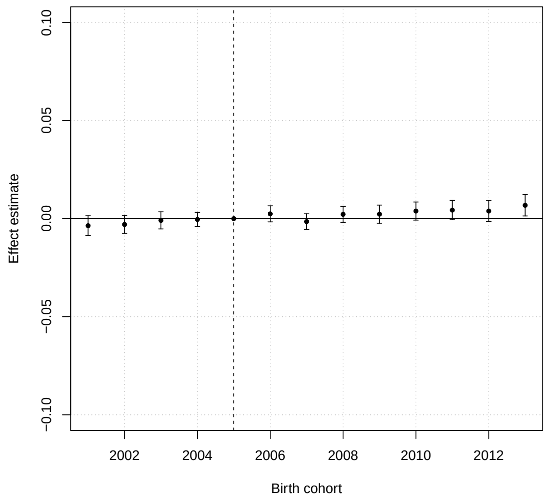
Notes: The graphs show event-study estimates for 6-8 year old children for infections (panel (a)), ear diseases (panel (b)) and respiratory diseases (panel (c)). The estimates are based on Equation (4.A.2).
Source: KBV 2009–2019, own calculations.

Figure 4.A.4: Event study: Other diseases

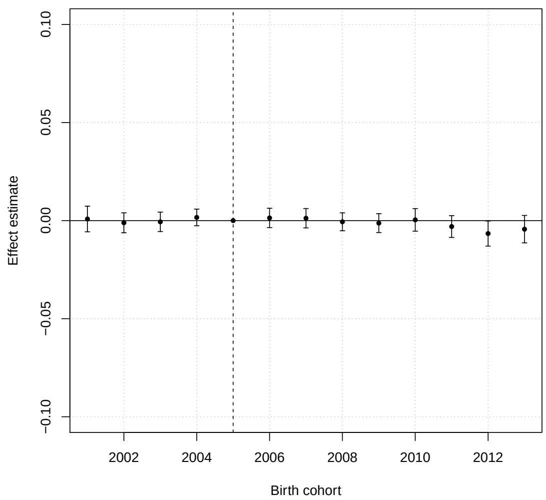
(a) Mental disorders



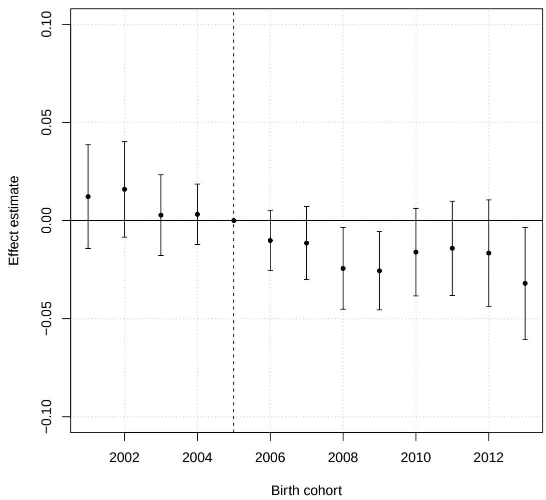
(b) Obesity



(c) Injury



(d) Vision problems

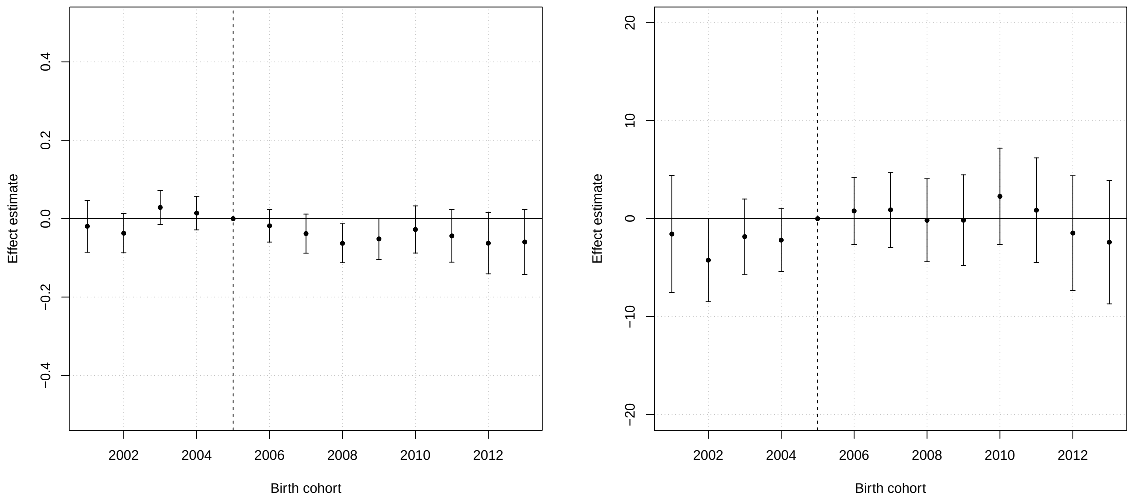


Notes: The graphs show event-study estimates for 6-8 year old children for mental disorders (panel (a)), obesity (panel (b)), injury (panel (c)) and vision problems (panel (d)). The estimates are based on Equation (4.A.2).

Source: KBV 2009–2019, own calculations.

Figure 4.A.5: Event study: Healthcare consumption and costs

(a) Treatment cases (b) Healthcare costs



Notes: The graphs show event-study estimates for 6-8 year old children for treatment cases (panel (a)) and healthcare costs (panel (b)). The estimates are based on Equation (4.A.2).
Source: KBV 2009–2019, own calculations.

Table 4.A.4: Generalized DiD Results: Squared term *cc*

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases				
Infections	0.011***	0.009***	0.001	−0.002
	(0.003)	(0.003)	(0.002)	(0.001)
Pre-Treatment Mean	1.394	1	0.777	0.665
Ear diseases	0.001	0.001	0.0005	0.0001
	(0.002)	(0.002)	(0.001)	(0.0005)
Pre-Treatment Mean	0.583	0.84	0.454	0.284
Respiratory diseases	0.017***	0.001	−0.003	−0.006**
	(0.004)	(0.005)	(0.002)	(0.002)
Pre-Treatment Mean	2.854	2.653	1.852	1.583
Non-communicable diseases				
Mental diseases	−0.0001	−0.0002	−0.001*	−0.002***
	(0.001)	(0.001)	(0.0005)	(0.0004)
Pre-Treatment Mean	0.177	0.37	0.329	0.275
Obesity	−0.00002	0.0002	0.0002	−0.0004*
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Pre-Treatment Mean	0.014	0.024	0.033	0.051
Injury	0.002***	−0.001*	−0.0003	0.0001
	(0.0004)	(0.0004)	(0.0003)	(0.0004)
Pre-Treatment Mean	0.216	0.19	0.189	0.239
Vision problems	0.0003	0.001+	−0.001+	−0.001+
	(0.001)	(0.001)	(0.0004)	(0.0004)
Pre-Treatment Mean	0.34	0.381	0.341	0.322
Healthcare consumption				
Treatment cases	0.005	−0.016**	0.001	0.005
	(0.007)	(0.006)	(0.007)	(0.006)
Pre-Treatment Mean	6.331	6.135	5.279	4.911
Healthcare costs	1.780***	0.154	0.353	−0.052
	(0.307)	(0.333)	(0.260)	(0.437)
Pre-Treatment Mean	319.964	287.281	244.466	249.115
Control for age + gender	yes	yes	yes	yes
Control for swine flu incidence	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	14,117,183	13,979,566	10,605,784

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.1. The term cc^2 is added. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.

Source: KBV 2009–2019, own calculations.

Table 4.A.5: DiD Results: Without Phase-in dummy

	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases			
Infections	−0.013	−0.017*	−0.007
	(0.012)	(0.008)	(0.007)
Pre-Treatment Mean	1.003	0.78	0.665
Ear diseases	0.005	−0.006	0.002
	(0.008)	(0.004)	(0.003)
Pre-Treatment Mean	0.929	0.461	0.289
Respiratory diseases	−0.025	−0.043**	−0.016
	(0.020)	(0.015)	(0.014)
Pre-Treatment Mean	2.951	1.951	1.634
Non-communicable diseases			
Mental disorders	0.007*	0.001	0.001
	(0.003)	(0.003)	(0.003)
Pre-Treatment Mean	0.881	0.959	0.975
Obesity	0.001	0.002**	0.002*
	(0.001)	(0.001)	(0.001)
Pre-Treatment Mean	0.025	0.032	0.05
Injury	−0.001	−0.001	−0.002
	(0.002)	(0.001)	(0.002)
Pre-Treatment Mean	0.199	0.2	0.248
Vision problems	−0.002	−0.0002	0.002
	(0.003)	(0.002)	(0.002)
Pre-Treatment Mean	0.374	0.334	0.32
Healthcare consumption			
Treatment cases	−0.062**	−0.048*	0.011
	(0.023)	(0.023)	(0.025)
Pre-Treatment Mean	6.439	5.35	4.936
Healthcare costs	−0.319	1.664	6.231*
	(1.302)	(1.730)	(2.569)
Pre-Treatment Mean	304.959	247.604	247.59
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,235,062	7,903,346	5,990,518

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.A.1 excluding the phase-in dummy. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on the respective disease.
Source: KBV 2009–2019, own calculations.

Table 4.A.6: Placebo Regression (generalized DiD): Diabetes

	Age: 1-10	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Infections	−0.00002	0.00003	−0.00001	−0.00005	−0.0001
	(0.00002)	(0.00003)	(0.00003)	(0.00004)	(0.00004)
Pre-Treatment Mean	0.002	0.001	0.001	0.002	0.003
Control for age + gender	yes	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes
Birth cohorts	2000-2014	2008-2014	2006-2014	2003-2011	2000-2009
Observations	54,152,607	8,522,318	14,117,165	13,979,538	10,605,769

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. Diabetes is coded as a dummy variable (indicating if a child had at least once per year a Diabetes diagnosis). The estimates are based on the specification in Equation 4.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on diabetes.
Source: KBV 2009–2019, own calculations.

Table 4.A.7: Placebo Regression (DiD): Diabetes

	Age: 3-10	Age: 3-5	Age: 6-8	Age: 9-10
Infections	−0.00001	0.0002	0.0001	−0.0002
	(0.0001)	(0.0002)	(0.0002)	(0.0002)
Pre-Treatment Mean	0.003	0.002	0.002	0.003
Control for age + gender	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2000-2011	2006-2011	2003-2011	2000-2009
Observations	21,215,384	5,235,058	7,903,335	5,990,512

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. Diabetes is coded as a dummy variable (indicating if a child had at least once per year a Diabetes diagnosis). The estimates are based on the specification in Equation 4.A.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on diabetes.
Source: KBV 2009–2019, own calculations.

Table 4.A.8: Generalized DiD Results: Exclusion of big cities

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases				
Infections	0.009***	0.002	−0.003***	−0.004***
	(0.002)	(0.001)	(0.001)	(0.001)
Pre-Treatment Mean	1.385	0.999	0.772	0.658
Ear diseases	0.003**	−0.001	−0.0001	−0.001+
	(0.001)	(0.001)	(0.0004)	(0.0003)
Pre-Treatment Mean	0.584	0.849	0.455	0.284
Respiratory diseases	0.014***	−0.001	−0.004*	−0.005***
	(0.003)	(0.003)	(0.002)	(0.002)
Pre-Treatment Mean	2.875	2.68	1.854	1.581
Non-communicable diseases				
Mental diseases	0.001***	0.0001	−0.001*	−0.001***
	(0.0003)	(0.0004)	(0.0003)	(0.0003)
Pre-Treatment Mean	0.177	0.372	0.327	0.273
Obesity	0.0001	0.0001+	0.0001	−0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Pre-Treatment Mean	0.014	0.023	0.032	0.05
Injury	0.0005*	−0.0003	0.00000	0.0001
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Pre-Treatment Mean	0.219	0.193	0.192	0.244
Vision problems	0.001***	0.0003	−0.0005+	−0.001**
	(0.0003)	(0.0004)	(0.0003)	(0.0002)
Pre-Treatment Mean	0.338	0.386	0.345	0.325
Healthcare consumption				
Treatment cases	0.007*	−0.014***	−0.005+	−0.007*
	(0.003)	(0.004)	(0.003)	(0.003)
Pre-Treatment Mean	6.328	6.16	5.28	4.901
Healthcare costs	1.391***	−0.065	−0.549**	−1.004***
	(0.175)	(0.204)	(0.183)	(0.252)
Pre-Treatment Mean	319.179	287.256	241.689	244.404
Control for age + gender	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009
Observations	7,153,668	11,998,338	12,020,533	9,197,199

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.1. Children residing in cities with more than 500,000 inhabitants are excluded. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.
Source: KBV 2009–2019, own calculations.

Table 4.A.9: DiD Results: Exclusion of big cities

	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases			
Infections	-0.011	-0.033^{***}	-0.017^{*}
	(0.011)	(0.010)	(0.008)
Pre-Treatment Mean	0.996	0.773	0.658
Ear diseases	0.005	-0.007	-0.0002
	(0.009)	(0.005)	(0.004)
Pre-Treatment Mean	0.938	0.461	0.288
Respiratory diseases	-0.034	-0.055^{*}	-0.029
	(0.021)	(0.022)	(0.018)
Pre-Treatment Mean	2.98	1.953	1.631
Non-communicable diseases			
Mental disorders	0.008^{*}	0.003	0.001
	(0.003)	(0.004)	(0.003)
Pre-Treatment Mean	0.383	0.303	0.261
Obesity	0.0004	0.002^{**}	0.002^{+}
	(0.001)	(0.001)	(0.001)
Pre-Treatment Mean	0.025	0.031	0.048
Injury	-0.0004	-0.001	-0.002
	(0.002)	(0.002)	(0.003)
Pre-Treatment Mean	0.202	0.204	0.252
Vision problems	-0.002	-0.001	0.002
	(0.003)	(0.003)	(0.003)
Pre-Treatment Mean	0.379	0.338	0.323
Healthcare consumption			
Treatment cases	-0.070^{**}	-0.070^{*}	0.00002
	(0.026)	(0.033)	(0.033)
Pre-Treatment Mean	6.466	5.35	4.925
Healthcare costs	0.083	3.431	8.307^{**}
	(1.410)	(2.295)	(3.096)
Pre-Treatment Mean	305.295	244.719	242.862
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	4,208,643	6,443,579	4,942,566

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.A.1. Children residing in cities with more than 500,000 inhabitants are excluded. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on the respective disease.
Source: KBV 2009–2019, own calculations.

Table 4.A.10: Infections: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	−0.012	−0.025**	−0.008
	(0.011)	(0.009)	(0.008)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	4,296,474	6,914,050	5,882,942
Exp. period: 2009–2012	−0.013	−0.020*	−0.006
	(0.012)	(0.010)	(0.009)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,181,917	7,811,109	5,913,604
Exp. period: 2009–2013	−0.011	−0.017+	−0.005
	(0.012)	(0.010)	(0.009)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	6,154,171	7,924,942	5,996,757
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on infections.
Source: KBV 2009–2019, own calculations.

Table 4.A.11: Ear diseases: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	0.014^{+}	0.003	0.004
	(0.007)	(0.005)	(0.003)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	4,296,469	6,914,046	5,882,937
Exp. period: 2009–2012	−0.007	−0.002	0.002
	(0.007)	(0.005)	(0.004)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,181,916	7,811,105	5,913,600
Exp. period: 2009–2013	−0.001	0.001	0.005
	(0.008)	(0.004)	(0.004)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	6,154,168	7,924,939	5,996,749
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on ear disease.
Source: KBV 2009–2019, own calculations.

Table 4.A.12: Respiratory diseases: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	-0.028	-0.051^{**}	-0.023
	(0.019)	(0.019)	(0.017)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	4,296,465	6,914,045	5,882,937
Exp. period: 2009–2012	-0.029	-0.039^{+}	-0.008
	(0.019)	(0.020)	(0.016)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,181,909	7,811,105	5,913,599
Exp. period: 2009–2013	-0.022	-0.037^{+}	-0.0004
	(0.020)	(0.019)	(0.015)
Observations	6,154,155	7,924,936	5,996,750
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on respiratory disease.

Source: KBV 2009–2019, own calculations.

Table 4.A.13: Mental disorders: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	0.005^{+}	0.003	0.002
	(0.003)	(0.003)	(0.003)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	4,296,474	6,914,045	5,882,936
Exp. period: 2009–2012	0.004	0.003	0.002
	(0.003)	(0.003)	(0.003)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,181,915	7,811,106	5,913,600
Exp. period: 2009–2013	0.004	0.001	0.002
	(0.003)	(0.003)	(0.003)
Observations	6,154,172	7,924,942	5,996,744
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Robust standard errors clustered on county-level in parentheses. The outcome variable is a dummy variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on mental disorders.

Source: KBV 2009–2019, own calculations.

Table 4.A.14: Obesity: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	0.001	0.002*	0.002*
	(0.001)	(0.001)	(0.001)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	0.0004	0.003***	0.002*
	(0.001)	(0.001)	(0.001)
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)	(0.001)
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 swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a dummy variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on obesity.
Source: KBV 2009–2019, own calculations.

Table 4.A.15: Injury: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	−0.002	−0.002	−0.004
	(0.002)	(0.002)	(0.002)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	4,296,474	6,914,045	5,882,936
Exp. period: 2009–2012	−0.002	−0.001	-0.006^{*}
	(0.002)	(0.002)	(0.002)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,181,915	7,811,106	5,913,600
Exp. period: 2009–2013	−0.001	−0.001	-0.004^{+}
	(0.002)	(0.002)	(0.002)
Observations	6,154,172	7,924,942	5,996,744
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on injuries.
Source: KBV 2009–2019, own calculations.

Table 4.A.16: Vision problems: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	-0.005^{+}	-0.002	0.002
	(0.003)	(0.003)	(0.002)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	7,715,602	12,448,997	10,605,640
Exp. period: 2009–2012	-0.003	-0.001	0.0004
	(0.003)	(0.003)	(0.003)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	9,241,260	13,979,440	10,605,640
Exp. period: 2009–2013	-0.003	-0.002	0.001
	(0.003)	(0.003)	(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 swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a dummy variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on vision problems.
Source: KBV 2009–2019, own calculations.

Table 4.A.17: Treatment cases: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	-0.062^{**}	-0.074^{*}	-0.006
	(0.022)	(0.029)	(0.029)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	4,296,473	6,914,043	5,882,939
Exp. period: 2009–2012	-0.078^{***}	-0.085^{**}	-0.018
	(0.023)	(0.027)	(0.029)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,181,916	7,811,107	5,913,601
Exp. period: 2009–2013	-0.086^{***}	-0.087^{**}	-0.008
	(0.024)	(0.027)	(0.028)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	6,154,169	7,924,935	5,996,752
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on treatment cases.

Source: KBV 2009–2019, own calculations.

Table 4.A.18: Healthcare costs: Different expansion period definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
Exp. period: 2008–2011	−1.267	0.314	6.901*
	(1.340)	(2.189)	(3.055)
Birth cohorts	2006-2010	2003-2010	2000-2009
Observations	4,296,470	6,914,049	5,882,939
Exp. period: 2009–2012	−1.009	−0.337	3.290
	(1.475)	(2.009)	(2.738)
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,181,913	7,811,105	5,913,602
Exp. period: 2009–2013	−1.597	0.066	4.715
	(1.444)	(2.198)	(3.051)
Birth cohorts	2006-2012	2003-2011	2000-2009
Observations	6,154,169	7,924,938	5,996,753
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the expansion period. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on healthcare costs.

Source: KBV 2009–2019, own calculations.

Table 4.A.19: Infections: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	−0.002	−0.011	−0.002
	(0.009)	(0.009)	(0.007)
Observations	9,241,248	13,979,422	10,605,626
upper 40 vs. lower 40%	−0.008	−0.021*	−0.008
	(0.010)	(0.009)	(0.008)
Observations	7,162,809	10,828,710	8,212,083
upper 35 vs. lower 35%	−0.010	−0.025**	−0.009
	(0.011)	(0.009)	(0.008)
Observations	6,107,240	9,224,953	6,992,512
upper 25 vs. lower 25%	−0.019	−0.030**	−0.011
	(0.014)	(0.011)	(0.010)
Observations	4,085,745	6,147,066	4,646,346
upper 20 vs. lower 20%	−0.018	−0.032**	−0.014
	(0.014)	(0.011)	(0.011)
Observations	3,406,869	5,105,933	3,849,284
percentage change	−0.001	−0.003*	−0.001
	(0.002)	(0.001)	(0.001)
Observations	9,241,248	13,979,422	10,605,626
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the $Treat$-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on infections.
Source: KBV 2009–2019, own calculations.

Table 4.A.20: Ear diseases: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	0.006	−0.001	0.001
	(0.006)	(0.004)	(0.003)
Observations	9,241,241	13,979,415	10,605,625
upper 40 vs. lower 40%	0.007	−0.004	−0.002
	(0.007)	(0.005)	(0.003)
Observations	7,162,812	10,828,713	8,212,080
upper 35 vs. lower 35%	0.005	−0.006	−0.001
	(0.007)	(0.005)	(0.003)
Observations	6,107,238	9,224,951	6,992,508
upper 25 vs. lower 25%	0.006	−0.003	0.004
	(0.009)	(0.005)	(0.004)
Observations	4,085,741	6,147,065	4,646,343
upper 20 vs. lower 20%	0.002	−0.005	0.004
	(0.009)	(0.006)	(0.004)
Observations	3,406,865	5,105,931	3,849,280
percentage change	0.001	−0.0003	0.0004
	(0.001)	(0.001)	(0.0004)
Observations	9,241,254	13,979,421	10,605,627
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the $Treat$-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on ear diseases.

Source: KBV 2009–2019, own calculations.

Table 4.A.21: Respiratory diseases: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	−0.020	−0.034*	−0.018
	(0.014)	(0.015)	(0.013)
Observations	9,241,241	13,979,415	10,605,625
upper 40 vs. lower 40%	−0.015	−0.039*	−0.022
	(0.016)	(0.017)	(0.014)
Observations	7,162,802	10,828,708	8,212,081
upper 35 vs. lower 35%	−0.017	−0.046*	−0.024
	(0.018)	(0.019)	(0.015)
Observations	6,107,230	9,224,947	6,992,509
upper 25 vs. lower 25%	−0.033	−0.050*	−0.024
	(0.023)	(0.024)	(0.019)
Observations	4,085,736	6,147,064	4,646,344
upper 20 vs. lower 20%	−0.031	−0.054*	−0.029
	(0.025)	(0.026)	(0.020)
Observations	3,406,862	5,105,930	3,849,280
percentage change	−0.003	−0.006*	−0.003
	(0.003)	(0.003)	(0.002)
Observations	9,241,241	13,979,415	10,605,625
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the $Treat$-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on respiratory diseases.

Source: KBV 2009–2019, own calculations.

Table 4.A.22: Mental disorders: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	0.04^{+}	0.002	0.001
	(0.002)	(0.002)	(0.002)
Observations	9,241,246	13,979,423	10,605,626
upper 40 vs. lower 40%	0.004^{+}	0.001	-0.0001
	(0.002)	(0.003)	(0.003)
Observations	7,162,807	10,828,712	8,212,079
upper 35 vs. lower 35%	0.006^{**}	0.001	0.0003
	(0.003)	(0.003)	(0.003)
Observations	6,107,233	9,224,949	6,992,509
upper 25 vs. lower 25%	0.007^{+}	0.004	0.003
	(0.004)	(0.004)	(0.004)
Observations	4,085,740	6,147,065	4,646,343
upper 20 vs. lower 20%	0.007^{+}	0.004	0.003
	(0.004)	(0.004)	(0.004)
Observations	3,406,864	5,105,931	3,849,280
percentage change	0.002^{+}	0.002	0.003
	(0.001)	(0.002)	(0.002)
Observations	9,241,246	13,979,423	10,605,626
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a dummy variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the *Treat*-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on mental disorders.

Source: KBV 2009–2019, own calculations.

Table 4.A.23: Obesity: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	0.0002	0.001*	0.001
	(0.0005)	(0.001)	(0.002)
Observations	9,241,244	13,979,425	10,605,619
upper 40 vs. lower 40%	0.0004	0.002*	0.002^{+}
	(0.001)	(0.001)	(0.001)
Observations	7,162,804	10,828,715	8,212,074
upper 35 vs. lower 35%	0.0005	0.002**	0.002^{+}
	(0.001)	(0.001)	(0.001)
Observations	6,107,232	9,224,952	6,992,502
upper 25 vs. lower 25%	0.001	0.003***	0.004**
	(0.001)	(0.001)	(0.001)
Observations	4,085,738	6,147,066	4,646,343
upper 20 vs. lower 20%	0.001	0.003***	0.004**
	(0.001)	(0.001)	(0.001)
Observations	3,406,863	5,105,937	3,849,280
percentage change	0.0002	0.0005*	0.0003
	(0.0002)	(0.0002)	(0.0003)
Observations	9,241,244	13,979,425	10,605,619
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a dummy variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the *Treat*-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on obesity. *Source:* KBV 2009–2019, own calculations.

Table 4.A.24: Injury: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	−0.0001	−0.001	−0.002
	(0.001)	(0.001)	(0.002)
Observations	9,241,251	13,979,422	10,605,622
upper 40 vs. lower 40%	0.0004	−0.001	−0.002
	(0.002)	(0.002)	(0.002)
Observations	7,162,811	10,828,713	8,212,076
upper 35 vs. lower 35%	−0.0002	−0.002	−0.003
	(0.002)	(0.002)	(0.002)
Observations	6,107,237	9,224,951	6,992,505
upper 25 vs. lower 25%	−0.001	−0.002	−0.004
	(0.002)	(0.002)	(0.003)
Observations	4,085,744	6,147,064	4,646,344
upper 20 vs. lower 20%	−0.0004	−0.002	−0.006*
	(0.002)	(0.002)	(0.003)
Observations	3,406,868	5,105,933	3,849,283
percentage change	−0.0001	−0.0001	−0.0004
	(0.0002)	(0.0002)	(0.0003)
Observations	9,241,251	13,979,422	10,605,622
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the *Treat*-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on injuries. *Source:* KBV 2009–2019, own calculations.

Table 4.A.25: Vision problems: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	−0.0003	0.001	0.002
	(0.002)	(0.002)	(0.002)
Observations	9,241,251	13,979,427	10,605,630
upper 40 vs. lower 40%	−0.002	−0.002	0.0001
	(0.002)	(0.002)	(0.002)
Observations	7,162,810	10,828,721	8,212,084
upper 35 vs. lower 35%	−0.002	−0.001	0.001
	(0.003)	(0.003)	(0.002)
Observations	6,107,236	9,224,958	6,992,513
upper 25 vs. lower 25%	−0.001	0.0001	0.004
	(0.003)	(0.003)	(0.003)
Observations	4,085,740	6,147,071	4,646,348
upper 20 vs. lower 20%	−0.003	−0.001	0.003
	(0.004)	(0.004)	(0.003)
Observations	3,406,863	5,105,937	3,849,285
percentage change	−0.001	-0.002^{*}	−0.002
	(0.001)	(0.001)	(0.001)
Observations	9,241,251	13,979,427	10,605,630
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a dummy variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the *Treat*-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on vision problems.

Source: KBV 2009–2019, own calculations.

Table 4.A.26: Treatment cases: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	-0.038^{*}	-0.048^{*}	0.001
	(0.019)	(0.022)	(0.022)
Observations	9,241,251	13,979,423	10,605,628
upper 40 vs. lower 40%	-0.040^{+}	-0.062^{*}	-0.00003
	(0.021)	(0.025)	(0.025)
Observations	7,162,812	10,828,713	8,212,081
upper 35 vs. lower 35%	-0.066^{**}	-0.082^{**}	-0.007
	(0.023)	(0.027)	(0.027)
Observations	6,107,240	9,224,950	6,992,509
upper 25 vs. lower 25%	-0.068^{*}	-0.059^{+}	0.010
	(0.027)	(0.034)	(0.033)
Observations	4,085,743	6,147,062	4,646,343
upper 20 vs. lower 20%	-0.070^{*}	-0.078^{*}	-0.023
	(0.029)	(0.037)	(0.035)
Observations	3,406,866	5,105,931	3,849,280
percentage change	-0.008^{*}	-0.010^{**}	-0.001
	(0.003)	(0.004)	(0.004)
Observations	9,241,251	13,979,423	10,605,628
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the $Treat$-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on treatment cases.

Source: KBV 2009–2019, own calculations.

Table 4.A.27: Healthcare costs: Different treatment definitions (DiD)

	Age: 3-5	Age: 6-8	Age: 9-10
upper 50 vs. lower 50%	0.586	0.764	3.878^{+}
	(1.040)	(1.595)	(2.229)
Observations	9,241,249	13,979,424	10,605,629
upper 40 vs. lower 40%	0.351	0.819	4.761^{+}
	(1.151)	(1.820)	(2.569)
Observations	7,162,807	10,828,712	8,212,082
upper 35 vs. lower 35%	−0.080	1.075	5.257^{+}
	(1.249)	(1.956)	(2.829)
Observations	6,107,235	9,224,949	6,992,509
upper 25 vs. lower 25%	−0.604	3.479	10.217^{**}
	(1.404)	(2.430)	(3.239)
Observations	4,085,742	6,147,066	4,646,343
upper 20 vs. lower 20%	−0.361	2.433	7.578^{*}
	(1.505)	(2.650)	(3.376)
Observations	3,406,865	5,105,932	3,849,280
percentage change	0.005	0.165	0.945^{*}
	(0.157)	(0.257)	(0.367)
Observations	9,241,249	13,979,424	10,605,629
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The outcome variable is a count variable. The estimates are based on the specification in Equation 4.A.1 with varying definitions of the $Treat$-Variable. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on healthcare costs.
Source: KBV 2009–2019, own calculations.

Table 4.A.28: Generalized DiD Results: Extensive/intensive margin

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases				
Infections	0.002***	0.001	−0.001*	−0.002***
	(0.0003)	(0.0004)	(0.0003)	(0.0003)
Sample Mean	0.63	0.534	0.456	0.404
Ear diseases	0.001***	−0.0001	−0.00000	−0.0003*
	(0.0003)	(0.0003)	(0.0002)	(0.0001)
Sample Mean	0.327	0.394	0.239	0.164
Respiratory diseases	0.002***	−0.0003	−0.001*	−0.001***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Sample Mean	0.81	0.772	0.648	0.585
Non-communicable diseases				
Mental disorders	0.001	−0.0005	−0.003	−0.006**
	(0.001)	(0.002)	(0.002)	(0.002)
Sample Mean	0.312	0.867	1.057	1.031
Obesity	0.0002*	0.0003*	−0.0001	−0.001***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Sample Mean	0.022	0.043	0.063	0.098
Injury	0.0003*	−0.0001	−0.00005	0.0002+
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Sample Mean	0.165	0.14	0.132	0.161
Vision problems	0.002**	−0.0004	−0.003**	−0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Sample Mean	0.592	0.816	0.791	0.709
Control for age + gender	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009
Observations	8,522,309	14,117,159	13,979,527	10,605,758

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): infections, ear diseases, respiratory diseases, injuries. The following variables are count variables (annual number of diagnoses): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.

Source: KBV 2009–2019, own calculations.

Table 4.A.29: DiD Results: Extensive/intensive margin

	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases			
Infections	-0.003	-0.009^{*}	-0.003
	(0.004)	(0.004)	(0.004)
Pre-Treatment Mean	0.536	0.457	0.404
Ear diseases	-0.0004	-0.003	-0.0004
	(0.002)	(0.002)	(0.001)
Pre-Treatment Mean	0.424	0.244	0.168
Respiratory diseases	-0.004^{+}	-0.009^{**}	-0.003
	(0.002)	(0.003)	(0.003)
Pre-Treatment Mean	0.798	0.663	0.594
Non-communicable diseases			
Mental disorders	0.025^{*}	0.029^{+}	0.025
	(0.010)	(0.017)	(0.017)
Pre-Treatment Mean	0.881	0.959	0.975
Obesity	0.001	0.003	0.001
	(0.001)	(0.002)	(0.003)
Pre-Treatment Mean	0.043	0.061	0.095
Injury	-0.001	-0.0001	-0.001
	(0.001)	(0.001)	(0.001)
Pre-Treatment Mean	0.146	0.14	0.166
Vision problems	-0.009	-0.021^{*}	-0.015^{+}
	(0.009)	(0.010)	(0.009)
Pre-Treatment Mean	0.821	0.78	0.704
Control for age + gender	yes	yes	yes
Control for swine flu incidence	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes
Birth cohorts	2006-2011	2003-2011	2000-2009
Observations	5,235,062	7,903,346	5,990,518

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): infections, ear diseases, respiratory diseases, injuries. The following variables are count variables (annual number of diagnoses): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.A.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a living in a fast-expanding county and being born after the reform on the respective disease.
Source: KBV 2009–2019, own calculations.

4.A.6 Detailed diagnoses

In this section, I provide results for more narrowly defined diagnoses (3- and 4-digit levels of ICD-10 codes).

Infections. Within infections, intestinal infectious diseases are responsible for about 29% of infections of 1–2-year-old children. Herein other gastroenteritis and colitis of infectious and unspecified origin (e.g. "abdominal influenza") account for about 81%. Intestinal infections and, therein, gastroenteritis are the only subgroups among the studied subgroups for which I can depict significant increases for 1–2-year-old children and decreases for elementary school-aged children (Appendix Table 4.A.30). Namely, a 10 percentage point increase in the coverage rate leads to an 11% and 9% increase in intestinal infections and gastroenteritis diagnoses for 1–2-year-old children. The reductions at age 9–10 years amount to 10% for intestinal infections and 12% for gastroenteritis.

Ear diseases. For the aggregated set of ear diseases, I find significant increases at age 1–2 years but no sizable changes in 3–10-year-old children. However, the more detailed analysis reveals that the reform led to a significant increase in middle ear infections (diseases of middle ear and mastoid) at age 1–2 years and decreases at older ages. The increase at age 1–2 years is most evident for the nonsuppurative and otitis media subgroup (Appendix Table 4.A.30).

Respiratory diseases. Increases in respiratory diseases at age 1–2 years are particularly pronounced for acute upper respiratory infections (herein acute upper respiratory infections of multiple and unspecified sites), other acute lower respiratory infections (herein acute bronchitis), and other diseases of the upper respiratory tract. The decrease in respiratory diseases for elementary school-age children can be mainly attributed to other acute lower respiratory infections (herein acute bronchitis) and other diseases of the upper respiratory tract (herein allergic rhinitis), which account combined for one-third of all respiratory conditions for 9–10-year-old children (Appendix Table 4.A.31). Within the group of chronic lower respiratory diseases, there is a significant increase in asthma at age 1–2 years and a significant decrease at age 6–8 years.

Non-communicable diseases. The analysis of the effects of the expansion on the aggregated set of mental disorders does not reveal clear effects. However, when looking into frequent subgroups within the generalized DiD framework, I find evidence that the expansion increases the prevalence of disorders of psychological development and decreases the prevalence of behavioral and emotional disorders, with onset usually occurring in childhood and adolescence at age 1–2 years. In addition, I depict a significant decrease in behavioral and emotional disorders at elementary school age, with onset usually occurring in childhood and adolescence. Similarly, despite null effects on the aggregated set of vision problems, the detailed analysis provides evidence for an increase in the prevalence of disorders of conjunctiva (herein conjunctivitis) at age 1–2 years and a decrease at elementary school age (Appendix Table 4.A.32).[51]

[51]Note, conjunctivitis is mainly caused by viruses that likely spread in daycare centers. The results on conjunctivitis are in line with the findings on communicable diseases.

Table 4.A.30: Detailed diagnoses (infections and ear diseases)

	Age: 1-10	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Infections					
Intestinal infectious diseases (A00-A09)	-0.0002	0.003^{***}	0.001^{+}	-0.001^{*}	-0.001^{***}
	(0.0002)	(0.001)	(0.0004)	(0.0002)	(0.0002)
q-value	0.452	0.004	0.209	0.040	0.001
Pre-Treatment Mean	0.159	0.272	0.182	0.122	0.103
Other gastroenteritis and colitis of	-0.0002	0.002^{**}	0.0004	-0.001^{**}	-0.001^{***}
infectious and unspecified origin (A09)	(0.0002)	(0.001)	(0.0004)	(0.0003)	(0.0002)
q-value	0.635	0.046	0.724	0.066	0.001
Pre-Treatment Mean	0.129	0.219	0.149	0.101	0.085
Other viral diseases (B25-B34)	0.0004	0.002	-0.001	-0.0004	-0.001^{*}
	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0005)
q-value	0.559	0.150	0.667	0.432	0.086
Pre-Treatment Mean	0.215	0.356	0.264	0.164	0.128
Viral infection of unspecified site (B34)	0.0005	0.002	-0.0005	-0.0004	-0.001^{+}
	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0005)
q-value	0.635	0.248	0.920	0.730	0.173
Pre-Treatment Mean	0.208	0.348	0.256	0.158	0.123
Other infectious diseases (B99-B99)	0.001^{**}	0.002^{*}	0.001	-0.0002	-0.0004
	(0.0003)	(0.001)	(0.001)	(0.0003)	(0.0003)
q-value	0.015	0.067	0.447	0.490	0.240
Pre-Treatment Mean	0.146	0.271	0.181	0.106	0.075
Other and unspecified infectious diseases (B99)	0.001^{**}	0.002^{*}	0.001	-0.0002	-0.0004
	(0.0003)	(0.001)	(0.001)	(0.0003)	(0.0003)
q-value	0.026	0.109	0.724	0.730	0.302
Pre-Treatment Mean	0.146	0.271	0.181	0.106	0.075
Ear diseases					
Diseases of middle ear and mastoid (H65-H75)	0.001	0.002^{**}	-0.002^{**}	-0.001^{*}	-0.0004^{*}
	(0.0005)	(0.001)	(0.001)	(0.0003)	(0.0002)
q-value	0.363	0.006	0.112	0.109	0.075
Pre-Treatment Mean	0.372	0.475	0.59	0.275	0.152
Suppurative and unspecified otitis media (H66)	-0.0002	0.0004	-0.002^{***}	0.0001	0.00004
	(0.0002)	(0.001)	(0.0004)	(0.0002)	(0.0001)
q-value	0.635	0.590	0.004	0.730	0.823
Pre-Treatment Mean	0.17	0.271	0.255	0.115	0.068
Nonsuppurative otitis media (H65)	0.001^{+}	0.002^{**}	-0.0005	-0.0004^{*}	-0.0002^{+}
	(0.0003)	(0.001)	(0.0005)	(0.0002)	(0.0001)
q-value	0.316	0.036	0.724	0.116	0.208
Pre-Treatment Mean	0.165	0.193	0.286	0.116	0.055
Other disorders of ear (H90-H95)	0.001^{**}	0.0002	0.001^{*}	0.001^{*}	-0.0001
	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0001)
q-value	0.015	0.384	0.121	0.040	0.376
Pre-Treatment Mean	0.09	0.048	0.134	0.088	0.062
Other hearing loss (H91)	0.00003	-0.00001	0.0002	0.00003	-0.0001
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
q-value	0.834	0.927	0.807	0.845	0.723
Pre-Treatment Mean	0.023	0.009	0.037	0.023	0.016
Control for age + gender	yes	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes
Birth cohorts	2000-2014	2008-2014	2006-2014	2003-2011	2000-2009
Observations	51,857,093	7,369,329	14,118,601	13,982,062	10,608,646

Notes: $^{+}$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. The estimates are based on the specification in Equation 4.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.
Source: KBV 2009–2019, own calculations.

Table 4.A.31: Detailed diagnoses (respiratory diseases)

	Age: 1-10	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Respiratory diseases					
Acute upper respiratory infections (J00-J06)	−0.0004	0.005**	−0.001	0.001	−0.001+
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
q-value	0.705	0.006	0.667	0.256	0.121
Pre-Treatment Mean	0.998	1.422	1.225	0.824	0.664
Acute nasopharyngitis (common cold, J00)	−0.001*	0.0001	0.001	0.0005	−0.0002
	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0003)
q-value	0.160	0.927	0.724	0.506	0.723
Pre-Treatment Mean	0.145	0.254	0.179	0.108	0.082
Acute pharyngitis (J02)	−0.0002	−0.0003	−0.00000	0.0002	−0.0001
	(0.0003)	(0.001)	(0.0005)	(0.0004)	(0.0003)
q-value	0.635	0.783	1	0.730	0.823
Pre-Treatment Mean	0.12	0.139	0.137	0.109	0.099
Acute tonsillitis (J03)	−0.0003	0.0005	−0.0002	0.001*	0.0003
	(0.0003)	(0.001)	(0.0004)	(0.0003)	(0.0002)
q-value	0.635	0.495	0.920	0.080	0.271
Pre-Treatment Mean	0.171	0.152	0.222	0.066	0.123
Acute laryngitis, tracheitis (J04)	0.0001	0.0003	0.001	0.00004	−0.0003+
	(0.0002)	(0.0004)	(0.0004)	(0.0002)	(0.0001)
q-value	0.635	0.521	0.724	0.905	0.173
Pre-Treatment Mean	0.054	0.079	0.068	0.043	0.036
Acute upper respiratory infections of	−0.001	0.004**	−0.001	−0.0001	−0.001*
multiple and unspecified sites (J06)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
q-value	0.635	0.046	0.795	0.932	0.057
Pre-Treatment Mean	0.639	1.018	0.794	0.493	0.39
Other acute lower respiratory infections (J20-J22)	−0.001	0.005***	0.001	−0.001	−0.001**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)
q-value	0.498	0.000	0.667	0.225	0.017
Pre-Treatment Mean	0.275	0.462	0.363	0.199	0.144
Acute bronchitis (J20)	−0.0002	0.004***	0.0004	−0.001	−0.001**
	(0.001)	(0.001)	(0.001)	(0.0005)	(0.0004)
q-value	0.830	0.001	0.920	0.380	0.047
Pre-Treatment Mean	0.248	0.415	0.328	0.179	0.129
Other diseases of upper respiratory tract (J30-J39)	−0.0002	0.003**	−0.001	−0.002**	−0.002***
	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)
q-value	0.718	0.018	0.447	0.019	0.001
Pre-Treatment Mean	0.406	0.355	0.495	0.378	0.361
Vasomotor and allergic rhinitis (J30)	−0.0001	0.0005	−0.0002	−0.001**	−0.002***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
q-value	0.834	0.326	0.884	0.046	0.001
Pre-Treatment Mean	0.121	0.022	0.069	0.151	0.215
Chronic rhinitis, nasopharyngitis and pharyngitis (J31)	−0.0001	0.002	0.0001	0.0002	0.0002
	(0.0004)	(0.001)	(0.001)	(0.0003)	(0.0002)
q-value	0.838	0.248	0.968	0.730	0.436
Pre-Treatment Mean	0.101	0.189	0.119	0.074	0.055
Chronic lower respiratory diseases (J40-J47)	−0.001	0.002+	−0.001	−0.003**	−0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
q-value	0.452	0.108	0.668	0.019	0.503
Pre-Treatment Mean	0.327	0.34	0.351	0.312	0.309
Bronchitis, not specified as acute or chronic (J40)	0.0001	−0.001	−0.0005	0.0004	0.001
	(0.0003)	(0.001)	(0.002)	(0.001)	(0.001)
q-value	0.834	0.582	0.948	0.730	0.302
Pre-Treatment Mean	0.1	0.178	0.125	0.07	0.057
Asthma (J45)	−0.001	0.002*	−0.0003	−0.002***	−0.001
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)
q-value	0.588	0.109	0.948	0.008	0.271
Pre-Treatment Mean	0.189	0.118	0.175	0.211	0.231
Other diseases of the respiratory system (J95-J99)	0.0002	0.002	0.001	0.0004	−0.0002
	(0.0005)	(0.001)	(0.001)	(0.0003)	(0.0003)
q-value	0.705	0.175	0.447	0.256	0.503
Pre-Treatment Mean	0.132	0.231	0.169	0.095	0.069
Other respiratory disorders (J98)	0.0002	0.002	0.001	0.0005	−0.0001
	(0.0005)	(0.001)	(0.001)	(0.0003)	(0.0003)
q-value	0.830	0.272	0.849	0.421	0.796
Pre-Treatment Mean	0.13	0.229	0.167	0.094	0.068
Control for age + gender	yes	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes
Birth cohorts	2000-2014	2008-2014	2006-2014	2003-2011	2000-2009
Observations	51,857,121	7,369,332	14,118,563	13,982,067	10,608,649

Notes: +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. The estimates are based on the specification in equation 4.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.
Source: KBV 2009–2019, own calculations.

Table 4.A.32: Detailed diagnoses (mental disorders and vision problems)

	Age: 1-10	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Mental disorders					
Disorders of psychological development (F80-F89)	-0.0001	0.001^{**}	0.0002	-0.0005^{+}	-0.001^{*}
	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
q-value	0.718	0.044	0.668	0.136	0.033
Pre-Treatment Mean	0.22	0.116	0.305	0.234	0.154
Specific developmental disorders of	0.00001	0.0004^{+}	0.0002	-0.0003	-0.0003
speech and language (F80)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0002)
q-value	0.977	0.143	0.884	0.421	0.271
Pre-Treatment Mean	0.163	0.056	0.261	0.175	0.085
Specific developmental disorder of motor function (F82)	-0.0002	0.001^{*}	-0.0004^{+}	-0.0003^{+}	-0.0001
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0001)
q-value	0.508	0.109	0.439	0.305	0.796
Pre-Treatment Mean	0.05	0.045	0.059	0.054	0.033
Mixed specific developmental disorders (F83)	-0.0001	0.0002^{*}	0.00001	-0.00001	-0.00004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
q-value	0.635	0.123	0.990	0.932	0.823
Pre-Treatment Mean	0.022	0.008	0.024	0.029	0.021
Unspecified disorder of psychological development (F89)	0.0001	0.00004	-0.00001	0.00005	-0.00004
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)
q-value	0.635	0.927	0.990	0.845	0.823
Pre-Treatment Mean	0.029	0.022	0.034	0.032	0.023
Behavioural and emotional disorders with onset usually	-0.001^{***}	-0.001^{*}	-0.0005^{+}	-0.001^{*}	-0.001^{***}
occurring in childhood and adolescence (F90-F98)	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
q-value	0.000	0.000	0.209	0.040	0.001
Pre-Treatment Mean	0.112	0.036	0.092	0.135	0.154
Other behavioural and emotional disorders with onset	-0.0004^{***}	-0.001^{*}	-0.0001	-0.0003^{*}	-0.001^{***}
usually occurring in childhood and adolescence (F98)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
q-value	0.006	0.046	0.884	0.066	0.001
Pre-Treatment Mean	0.046	0.014	0.041	0.058	0.057
Vision problems					
Disorders of conjunctiva (H10-H13)	0.0004^{**}	0.002^{***}	-0.0005^{*}	-0.0002	-0.0003^{**}
	(0.0001)	(0.0003)	(0.0002)	(0.0001)	(0.0001)
q-value	0.015	0.000	0.184	0.244	0.597
Pre-Treatment Mean	0.143	0.418	0.195	0.096	0.071
Conjunctivitis (H10)	0.0004^{**}	0.002^{***}	-0.001^{*}	-0.0001	-0.0003^{**}
	(0.0001)	(0.0003)	(0.0002)	(0.0001)	(0.0001)
q-value	0.027	0.000	0.411	0.506	0.045
Pre-Treatment Mean	0.139	0.242	0.19	0.092	0.008
Diseases of the eye and adnexa (H53-H54)	-0.0002	-0.0003	0.0004	-0.0002	-0.0001
	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0003)
q-value	0.452	0.927	0.333	0.334	0.631
Pre-Treatment Mean	0.073	0.025	0.078	0.084	0.081
Visual disturbances (H53)	-0.0002	-0.0003	0.0003	-0.0001	0.00003
	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0002)
q-value	0.635	0.409	0.724	0.730	0.946
Pre-Treatment Mean	0.065	0.023	0.069	0.076	0.074
Visual impairment including blindness (H54)	0.00001	0.00002	0.00003	-0.00005	0.00000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
q-value	0.919	0.927	0.968	0.828	0.976
Pre-Treatment Mean	0.011	0.003	0.014	0.011	0.011
Control for age + gender	yes	yes	yes	yes	yes
Control for swine flu incidence	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes
Birth cohorts	2000-2014	2008-2014	2006-2014	2003-2011	2000-2009
Observations	21,221,456	5,235,816	7,905,234	5,993,036	

Notes: $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Robust standard errors clustered on county-level are in parentheses. All variables are dummy variables. The estimates are based on the specification in Equation 4.1. Outliers are excluded, i.e. the top 0.00001% in terms of number of diagnoses. The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.
Source: KBV 2009–2019, own calculations.

4.A.7 Heterogeneity

Table 4.A.33: Results by gender

	Girls				Boys			
	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases								
Infections	0.008***	0.002	−0.002**	−0.004***	0.008***	0.001	−0.003**	−0.004***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Pre-Treatment Mean	1.387	1.008	0.821	0.718	1.412	0.998	0.736	0.614
Ear diseases	0.003***	−0.001	−0.0003	−0.001*	0.003**	−0.0005	0.0002	−0.0005
	(0.001)	(0.001)	(0.0005)	(0.0003)	(0.001)	(0.001)	(0.001)	(0.0004)
Pre-Treatment Mean	0.534	0.816	0.452	0.289	0.634	0.868	0.457	0.28
Respiratory diseases	0.015***	−0.001	−0.004*	−0.005***	0.016***	−0.002	−0.004*	−0.006***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Pre-Treatment Mean	2.679	2.513	1.738	1.473	3.045	2.8	1.968	1.693
Non-communicable diseases								
Mental disorders	0.001**	0.0001	−0.001+	−0.001**	0.001**	0.0001	−0.001**	−0.001***
	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0003)
Pre-Treatment Mean	0.162	0.323	0.262	0.217	0.193	0.418	0.395	0.332
Obesity	0.0001+	0.0002**	0.0001	−0.0005***	0.0001	0.00002	0.00002	−0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Pre-Treatment Mean	0.015	0.027	0.035	0.052	0.013	0.02	0.03	0.051
Injury	0.001**	−0.0003	0.00003	0.0003	0.0003	−0.0003	−0.0002	0.00004
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Pre-Treatment Mean	0.195	0.165	0.172	0.227	0.239	0.215	0.206	0.252
Vision problems	0.001***	0.001	−0.0003	−0.0004	0.001***	0.0002	−0.0005+	−0.001**
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0002)
Pre-Treatment Mean	0.331	0.376	0.342	0.33	0.351	0.388	0.341	0.314
Healthcare consumption								
Treatment cases	0.010**	−0.011**	−0.006*	−0.004	0.010**	−0.014***	−0.006*	−0.005+
	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Pre-Treatment Mean	6.129	5.888	5.058	4.769	6.573	6.402	5.507	5.055
Healthcare costs	1.384***	−0.004	−0.513**	−0.750**	1.572***	−0.199	−0.504*	−1.095***
	(0.167)	(0.184)	(0.160)	(0.242)	(0.184)	(0.220)	(0.213)	(0.315)
Pre-Treatment Mean	308.688	268.586	219.012	222.848	333.192	306.602	269.764	275.062
Control for age + gender	yes	yes	yes	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009	2008-2014	2006-2014	2003-2011	2000-2009
Observations	4,169,396	6,919,385	6,882,828	5,223,170	4,306,952	7,144,969	7,060,343	5,353,428

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.1 (separately for boys and girls). The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.
Source: KBV 2009–2019, own calculations.

Table 4.A.34: Results by household income on county level

	Top 30th percentile household income				Bottom 30th percentile household income			
	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases								
Infections	0.001	−0.002	−0.001	−0.004*	0.011**	0.0004	−0.004**	−0.004**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)
Pre-Treatment Mean	1.353	0.96	0.758	0.647	1.448	1.04	0.803	0.694
Ear diseases	0.002+	−0.002	0.0004	−0.001	0.0001	0.001	0.00004	−0.001+
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Pre-Treatment Mean	0.55	0.796	0.432	0.269	0.628	0.895	0.482	0.304
Respiratory diseases	0.012***	−0.004	−0.001	−0.005+	0.012*	0.003	−0.003	−0.005
	(0.003)	(0.004)	(0.002)	(0.003)	(0.005)	(0.006)	(0.003)	(0.003)
Pre-Treatment Mean	2.701	2.459	1.724	1.476	3.09	2.911	2.028	1.736
Non-communicable diseases								
Mental disorders	0.001	0.0001	−0.001+	−0.001+	0.001*	0.0003	−0.001	−0.001+
	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)
Pre-Treatment Mean	0.171	0.358	0.32	0.266	0.18	0.377	0.338	0.284
Obesity	0.0001	0.0001	0.0002	−0.0005*	0.00003	0.0002	0.0001	−0.0004*
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Pre-Treatment Mean	0.013	0.022	0.03	0.048	0.015	0.026	0.035	0.055
Injury	0.0003	0.00002	0.0001	0.0005	0.0003	−0.001	−0.0005+	−0.0005
	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
Pre-Treatment Mean	0.206	0.18	0.18	0.231	0.228	0.201	0.199	0.251
Vision problems	0.0003	−0.001	−0.001	−0.001*	0.001**	0.001+	0.0001	−0.0002
	(0.0005)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.001)	(0.001)	(0.0005)
Pre-Treatment Mean	0.344	0.373	0.332	0.317	0.341	0.391	0.353	0.326
Healthcare consumption								
Treatment cases	0.010+	−0.013*	−0.006	−0.003	0.004	−0.014*	−0.005	−0.008
	(0.005)	(0.006)	(0.004)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)
Pre-Treatment Mean	6.15	5.891	5.096	4.763	6.547	6.417	5.523	5.12
Healthcare costs	1.399***	−0.026	−0.306	−0.482	1.036***	−0.044	−0.232	−1.341**
	(0.273)	(0.330)	(0.245)	(0.345)	(0.232)	(0.317)	(0.325)	(0.510)
Pre-Treatment Mean	314.933	278.307	240.285	249.829	327.273	299.153	253.064	254.179
Control for age + gender	yes	yes	yes	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009	2008-2014	2006-2014	2003-2011	2000-2009
Observations	2,537,818	4,235,551	4,168,113	3,171,942	2,506,540	4,205,943	4,186,757	3,166,076

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.1 (separately for children from counties in the top 30 income percentile and children from counties in the bottom 30 income percentile). The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.
Source: KBV 2009–2019, own calculations.

Table 4.A.35: Results by share of migrants on county level

	Top 30th percentile share of migrants				Bottom 30th percentile share of migrants			
	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Communicable diseases								
Infections	0.005*	0.0004	−0.003	−0.006*	0.008***	0.001	−0.001	−0.002*
	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Pre-Treatment Mean	1.415	1.008	0.791	0.686	1.376	0.985	0.754	0.636
Ear diseases	0.005***	0.001	0.001	−0.0004	0.001	−0.003*	−0.001	−0.001*
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.0004)
Pre-Treatment Mean	0.561	0.793	0.44	0.28	0.605	0.874	0.466	0.29
Respiratory diseases	0.027***	0.016**	−0.005	−0.009**	0.009*	−0.006	−0.001	−0.002
	(0.004)	(0.006)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)
Pre-Treatment Mean	2.853	2.61	1.863	1.606	2.882	2.674	1.837	1.57
Non-communicable diseases								
Mental disorders	0.0004	0.0005	−0.002*	−0.002*	0.001	−0.001+	−0.001	−0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)	(0.0004)	(0.0005)
Pre-Treatment Mean	0.177	0.36	0.329	0.276	0.186	0.382	0.331	0.277
Obesity	0.0001	0.0001	0.0001	−0.001***	0.00004	0.0002+	−0.00002	−0.0001
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Pre-Treatment Mean	0.015	0.026	0.037	0.059	0.014	0.022	0.029	0.046
Injury	0.001*	0.0003	0.00002	0.001	−0.0000	−0.001+	−0.0002	−0.0001
	(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
Pre-Treatment Mean	0.209	0.177	0.171	0.219	0.219	0.201	0.204	0.256
Vision problems	0.001*	0.0004	−0.0004	−0.0002	0.001+	−0.00003	−0.0001	−0.0003
	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.001)	(0.0004)	(0.0004)
Pre-Treatment Mean	0.336	0.36	0.319	0.304	0.345	0.397	0.363	0.343
Healthcare consumption								
Treatment cases	0.030***	0.015+	−0.009*	−0.006	−0.002	−0.027***	−0.007+	−0.003
	(0.006)	(0.008)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Pre-Treatment Mean	6.237	5.983	5.186	4.864	6.475	6.267	5.37	4.972
Healthcare costs	1.818***	0.635	−0.818*	−1.648**	1.006***	−0.660*	−0.759**	−0.610+
	(0.264)	(0.415)	(0.316)	(0.575)	(0.213)	(0.255)	(0.251)	(0.314)
Pre-Treatment Mean	314.976	279.585	243.168	253.364	330.728	295.077	247.763	248.453
Control for age + gender	yes	yes	yes	yes	yes	yes	yes	yes
Control for KKZ + Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Birth cohorts	2008-2014	2006-2014	2003-2011	2000-2009	2008-2014	2006-2014	2003-2011	2000-2009
Observations	2,522,457	4,206,470	4,207,269	3,182,483	2,418,955	4,126,573	4,111,397	3,086,171

Notes: +p<0.1; *p<0.05; **p<0.01; ***p<0.001. Robust standard errors clustered on county-level in parentheses. The following variables are count variables: infections, ear diseases, respiratory diseases, injuries (annual number of diagnoses), treatment cases and costs. Costs are fee-adjusted. The following variables are dummy variables (indicating if a child had at least once per year a particular diagnosis): mental disorders, obesity, vision problems. The estimates are based on the specification in Equation 4.1 (separately for children from counties in the top 30 share of migrants percentile and children from counties in the bottom 30 share of migrants percentile). The coefficients show the effect of a one percentage point increase in the daycare coverage rate on the respective disease.
Source: KBV 2009–2019, own calculations.

4.A.8 Additional analysis: SOEP

Table 4.A.36: Additional analysis: Parental health

	Age: 1-2	Age: 3-5	Age: 6-8	Age: 9-10
Maternal health	-0.0754*	-0.00462	0.0419	-0.0245
	(0.0384)	(0.0345)	(0.0476)	(0.0573)
Observations	8823	8187	4825	2650
Mother: Number of doctor visits	0.201+	-0.119	0.119	0.0321
	(0.120)	(0.213)	(0.202)	(0.188)
Observations	7218	8180	4819	2634
Mother: Number of days missed at work due to sickness	4.087***	-0.447	1.498+	0.0430
	(1.203)	(1.164)	(0.883)	(1.745)
Observations	3493	4908	3151	982
Mother: Number of days missed at work due to child's sickness	1.278**	-1.264***	-0.629	0.537
	(0.400)	(0.371)	(0.428)	(0.549)
Observations	1357	2059	1142	271
Paternal health	-0.0640+	0.0267	0.183***	-0.110
	(0.0387)	(0.0416)	(0.0501)	(0.0673)
Observations	7150	6136	3967	2142
Father: Number of doctor visits	0.0675	-0.0318	-0.274	-0.152
	(0.113)	(0.177)	(0.172)	(0.202)
Observations	5414	6118	3958	2134
Father: Number of days missed at work due to sickness	0.222	-0.345	-1.419	-1.570
	(1.103)	(0.954)	(1.131)	(1.245)
Observations	3567	4510	3064	921
Father: Number of days missed at work due to child's sickness	0.702	0.119	-0.585	-1.017
	(0.498)	(0.275)	(0.577)	(0.969)
Observations	645	781	416	77

Notes: $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001. Standard errors in parentheses. The estimates are based on a simple OLS regression where "daycare attendance at age 1–2 years" is the explanatory variable. The set of control variables includes parental education, survey year, cohabitation status, birth order, parental labor force status, parental migration background, household income, parental age, child sex, federal state of residence, age of siblings and if all-day daycare/school is attended.
Source: SOEP, v37, own calculations.

CHAPTER 5

The Effects of an Increase in the Retirement Age on Health Care Costs – Evidence from Administrative Data[1]

5.1 Introduction

Aging populations and demographic change challenge the financial stability of public pension systems. Therefore, many countries reform their pension systems and prolong work lives to increase contributions and to reduce the number of benefit recipients. However, an increasing retirement age might have adverse effects in other areas of the welfare state, specifically for the healthcare system. Previous studies (e.g., Atalay and Barrett, 2014; Barschkett et al., 2022; Kuusi et al., 2020; Shai, 2018) have documented that a prolonged working life can have negative health effects for individuals.[2] Yet, so far there exists no clear evidence how these negative health consequences affect healthcare costs. To assess the overall fiscal effects of pension reforms, this information is crucial.

[1]This chapter is joint work with Johannes Geyer (DIW Berlin), Peter Haan (DIW Berlin and Freie Universität Berlin and Anna Hammerschmid (former DIW Berlin). We are grateful to the National Association of Statutory Health Insurance Physicians (Kassenärztliche Bundesvereinigung, KBV) for data access and for their excellent support. We also thank three anonymous referees and the editor of this issue of The European Journal of Health Economics for valuable comments and suggestions.

[2]There is also evidence for positive effects of retirement on health (e.g., Atalay and Barrett, 2014; Atalay et al., 2019; Belloni et al., 2016; Charles, 2004; Coe and Zamarro, 2011; Eibich, 2015; Gorry et al., 2018; Grip et al., 2012; Leimer and Van Ewijk, 2022), yet findings are ambiguous and very much depend on the setting under study (e.g., pension reform, healthcare system etc.).

This paper uses unique data that cover outpatient care and associated costs to quantify the healthcare costs of a sizable increase in the retirement age.[3] The data include the universe of individuals insured through the German public healthcare system (almost 90% of the German population) and comprise a ten-year observation period (2009 – 2018). In addition to the overall health cost effect of the pension reform, the data also allow us to quantify separate cost effects for different medical specialist categories.

We exploit a sizable cohort-specific pension reform that abolished an early retirement scheme for all women born after 1951. The reform provides a clean quasi-experimental setting as it induces a substantial discontinuity in retirement behavior for two adjacent cohorts. We use this discontinuity in a Difference-in-Differences (DiD) estimation. This framework accounts for cohort and seasonality effects and allows us to identify the causal effect of the pension reform on healthcare costs. Specifically, similar to previous studies (e.g., Barschkett et al., 2022; Schönberg and Ludsteck, 2014), we define a treatment group (women born between October 1951 and March 1952) and a control group (women born between October 1950 and March 1951) and compare the healthcare costs of these groups over time.

Our results show that outpatient care costs significantly increase by about 2.9% (about 16 euro per individual) in the age group directly affected by the increase in the retirement age (60–62). Moreover, we also find expectation effects for women at the age of 59 and indirect post-employment effects for women between 63 and 65. We further show that the cost increase is mainly driven by utilization of the following specialist groups: Ophthalmologists, general practitioners (GPs), oral and maxillofacial surgery, neurology, orthopedics, and radiology. The absolute effect is largest for GPs (about 3.5 euro) and thereby contributes about 25% to the increase in the overall costs. While the effects are significant and meaningful on the individual level, we show that the increase in healthcare costs are modest relative to the positive fiscal effects of the pension reform. Specifically, we estimate an aggregate increase in the health costs of about 7.7 million euro for women aged 60-62 and born in 1952. The corresponding estimate of the net effects of the pension reform for the tax and transfer system including social security amounts to about 4 billion euro (Geyer et al., 2020).

Thus, from an aggregate perspective, our results of an increase in the healthcare costs do not provide strong evidence against an increase in the retirement age. However, the increase of costs on the individual level shows that positive fiscal effects of a longer working life can be counteracted by potential negative health consequences for

[3]Since the pension reform we consider mainly affected mental health, musculoskeletal diseases, and obesity healthcare costs related to outpatient care are – at least in the short run – of central importance Barschkett et al. (2022).

individuals. Moreover, our cost estimate focuses on the public healthcare costs and abstracts from individual disutility or other disadvantages due to worse health as well as other societal costs such as a decrease in labor productivity or an increase in sickness absence at work. For political decisions on retirement ages, such factors also need to be taken into account.

There exists a large literature on the health effects of retirement and pension reforms:[4] Some studies are based on survey data and explore effects of retirement on mental, physical or general health (e.g., Atalay and Barrett, 2014; Atalay et al., 2019; Belloni et al., 2016; Charles, 2004; Eibich, 2015; Etgeton and Hammerschmid, 2019; Gorry et al., 2018; Grip et al., 2012; Leimer and Van Ewijk, 2022). Others use administrative data and consider mortality (e.g., Brockmann et al., 2009; Fitzpatrick and Moore, 2018; Kuhn et al., 2019) or healthcare usage and diagnoses as outcome variables (e.g., Barschkett et al., 2022; Hagen, 2018; Horner and Cullen, 2016; Kuusi et al., 2020; Nielsen, 2019). The evidence of this literature is mixed and strongly depends on the pension reform[5] and the health outcomes[6] considered. Often broad health measures disguise effects of pension reforms on specific health outcomes. For example, using the same data source, Barschkett et al. (2022) show that the reform considered in this study specifically affects mental health, musculoskeletal diseases, and obesity. They find prolonging working life increases the prevalence of all mentioned diagnoses. The underlying reasons for the association between (mental) health and retirement may by mani-folded: Different stress-levels in and out of the labor force, changes in social contacts and mobility/movement are some examples.

Despite this sizable literature on health outcomes, there is only little evidence on the effects of pension reforms on public healthcare costs. Two examples are studies looking at pension reforms that delayed retirement with mixed evidence. Shai (2018) finds negative health effects of continued working and an increase in healthcare consumption in Israel. In contrast, Perdrix (2021) shows the opposite effect for France: she finds

[4]For a more detailed discussion, see e.g., Barschkett et al. (2022).

[5]The majority of previous studies on the link between health and retirement use age discontinuities in the retirement age to instrument the individual's retirement status (see van Ours and Picchio, 2020, for an overview of methodologies of previous studies). Only a few studies exploit direct variation from pension reforms (e.g., Barschkett et al., 2022; Bloemen et al., 2017; Charles, 2004; Etgeton and Hammerschmid, 2019; Grip et al., 2012; Kuhn et al., 2019).

[6]Health outcomes differ very much and range from self-assessed general health status to more specific self-assessed outcomes (e.g., cognitive abilities, mobility limitations, grip strength, hypertension, migraine, back pain) to mortality and specific diagnoses assessed by healthcare professionals (e.g., mental disorders, musculoskelatal diseases, cardiovascular diseases, obesity). Due to the wide range of outcomes as well as different assessment methods it is difficult to compare the effects and draw conclusions.

that later retirement leads to lower healthcare consumption. Associated with the lower number of doctor visits, she also finds lower expenditure.[7]

Our paper is structured as follows: In section 5.2, we describe the German pension and healthcare systems. Section 5.3 provides an overview on the data and section 5.4 explains the empirical strategy. In section 5.5, we describe the results and compare the additional costs for healthcare to the overall fiscal effects of the 1999 pension reform. Section 5.6 concludes.

5.2 Institutional background

In this section, we provide a brief overview on the relevant institutions of the German pension system and discuss the 1999 pension reform, which induced an exogenous increase in the early retirement age for women born after 1951. Moreover, we describe the German healthcare system.

5.2.1 Pension system

The German public pension system covers roughly 90% of the workforce.[8] Pension benefits account for about two-thirds of gross income of the elderly. The system is financed by a pay-as-you-go (PAYG) scheme and has a strong contributory link. The statutory pension age (SRA) was 65 for cohorts born before 1947. It is raised stepwisely to age 67 and fully phased in for all cohorts born in 1964 or later. For the 1951 cohort, the SRA was 65 and 5 months, for those born in 1952 it was 65 and 6 months. People qualify for this regular old-age pension after five years of pension contributions.

Women born before 1952 could retire before the SRA (with permanent deductions) at the age of 60 via the *pension for women*. The 1999 reform abolished this pathway to retirement for cohorts born after 1951. Effectively, the reform raised the early retirement age (ERA) for most women from 60 to 63, which implies an extension of the working life of three years. The eligibility criteria of the *pension for women* were: (i) at least 15 years of pension insurance contributions; and (ii) at least 10 years of

[7] Zhang et al. (2018) focus on private health expenditures and find for China that retirement increases healthcare utilization and yearly out of pocket expenditures for inpatient care as well as monthly out of pocket expenditures for self-treatment. For men, they also depict an increase in out-of-pocket inpatient costs.

[8] There are a few exemptions from compulsory insurance: civil servants have a separate tax-financed, non-contributory defined benefit scheme and most of the self-employed are not compulsorily insured (for a general description of the German pension system and the pension reform, see Barschkett et al., 2022; Börsch-Supan and Wilke, 2004).

pension insurance contributions after the age of 40. About 60% of all women born in 1951 were eligible for the old-age pension for women (Geyer and Welteke, 2021).[9]

5.2.2 Healthcare system

German residents are required to have health insurance.[10] About 90% of the population is insured in the public healthcare system.[11] People who opt out of the public system need to insure themselves in a private health insurance plan. Importantly, the insurance status is not affected by entry in retirement. Individuals with a public health insurance during the working life remain in this insurance during retirement.

Public health insurance is financed primarily through mandatory contributions by employers and employees[12], along with tax revenues. The public insurance offers insurance for non-contributing family members (family insurance). For individuals who receive unemployment benefits, the unemployment agency covers the contributions. For retirees, the pension insurance co-finances the contributions.

In Germany, publicly insured patients do not need to advance the costs of insured healthcare services. Instead, medical service providers settle their accounts via their regional Association of Statutory Health Insurance Physicians. Price and quantity parameters in the healthcare system are negotiated on a yearly basis by the National Association of Statutory Health Insurance Physicians and the National Association of Statutory Health Insurance Funds as well as their regional counterparts (see Appendix 5.A.1).

[9] Previous studies evaluate the labor market effects of the 1999 pension reform (Geyer et al., 2020; Geyer and Welteke, 2021). Based on data of the pension insurance Geyer and Welteke (2021) document that the labor market outcomes are very similar in the treatment and the control group before the age of 60. Moreover, they show that employment rates increased by about 15 percentage points (pre-reform mean 54%) and inactivity and unemployment increased by about 11 percentage points (pre-reform mean 12%). Further they point out that the reform caused women to stay longer in the current status, i.e., employed women continue working, unemployed women stay unemployed and inactive women remain inactive until reaching the new early retirement age. Thus, the negative health effects found by Barschkett et al. (2022) are mostly driven by women who stay longer in employment (e.g., due to increased stress-levels when working compared to being retired) and women who stay longer in unemployment (e.g., lower life-satisfaction due to delayed change of identity from unemployed to retired, Hetschko et al., 2014). In our data we are not able to differentiate the employment status of the women.

[10] For most information on the healthcare system in this section and additional details, see (BMG, 2020).

[11] There are a few exemptions from compulsory public insurance: e.g. people with an income above a certain threshold (5,213 euro monthly earnings in 2020), self-employed, and civil servants, are allowed to opt out of the public insurance.

[12] The overall contribution rate in 2020 was 14.6% of gross labor earnings, equally shared by employees and employers.

5.3 Data

We use administrative data covering the period from 2009 to 2018. The data stem from the database of claims of all publicly insured individuals in Germany as collected by the National Association of Statutory Health Insurance Physicians (KBV). For the analysis we use information on all insured women born between 1950–1952.[13] In addition to the group of women around the cutoff date of the pension reform (women born in late 1951 and in early 1952), we construct a control group consisting of women born late in 1950 and early in 1951.

The data include information for each patient about services and associated costs that medical specialists billed. For each patient the data contain yearly aggregated costs and costs that are specific to medical specialists.[14] In other words, each patient constitutes an entry for each year in the data set including information about the aggregated costs as well as the specific costs for each of the medical specialists.[15] The final data set includes about 500,000 women per birth cohort resulting in 1.5 million women overall. While the data includes detailed information on health outcomes and health costs, the data provides no information on important demographic variables such as education, employment status or income. Therefore, we cannot study the heterogeneous costs effects of the pension reform.

5.4 Empirical strategy

We estimate the effect of an increase in the retirement age on healthcare costs using a DiD estimation strategy. The medical literature (e.g., Boland et al., 2015; Doblhammer and Vaupel, 2001) documents that month of birth can affect health. Despite the set-up calling for an RDD approach, we prefer the DiD strategy as this allows us to account for seasonality. Specifically, following Schönberg and Ludsteck (2014) and Barschkett et al. (2022), we define a control group (women born between October 1950 and March 1951) and a treatment group (women born between October 1951 and March 1952). Women born between January and March are considered to be born after the cutoff. Thus, the

[13] We focus only on cohorts 1950–1952 since a major school reform affected many women born after 1952. Specifically, regional school reforms in West Germany raised compulsory schooling from 8 to 9 years. Four large West German federal states changed compulsory schooling within cohort 1953. The reform had positive effects on health outcomes (Kemptner et al., 2011).

[14] These costs are reported on the calendar quarter level in the original data. We aggregate the costs specific to medical specialists to the year level. Specialists not relevant for our research question (e.g., pediatricians) are not considered in this analysis.

[15] The sample is unbalanced as patients only appear if they received outpatient care at least once per year. Based on this information, we construct a balanced sample with yearly information on all publicly insured individuals. Costs for years without outpatient care are assumed to be zero.

interaction term between treatment group and being born after the cutoff estimates the effect of the pension reform. Importantly, the sample only includes individuals born between October 1951 and March 1952 as well as between October 1950 and March 1951, respectively. Thus, birth months between March and October are not included in the sample. This way, we avoid comparing birth months that are rather far away from the reform cutoff in January.

We account for correlation between observations of the same individual or individuals born in the same month, and use robust standard errors clustered by month of birth. In the subgroup analysis (costs by medical specialist), we additionally adjust the standard errors for multiple hypotheses testing using the Bonferroni-correction.

More formally, we estimate the following equation:

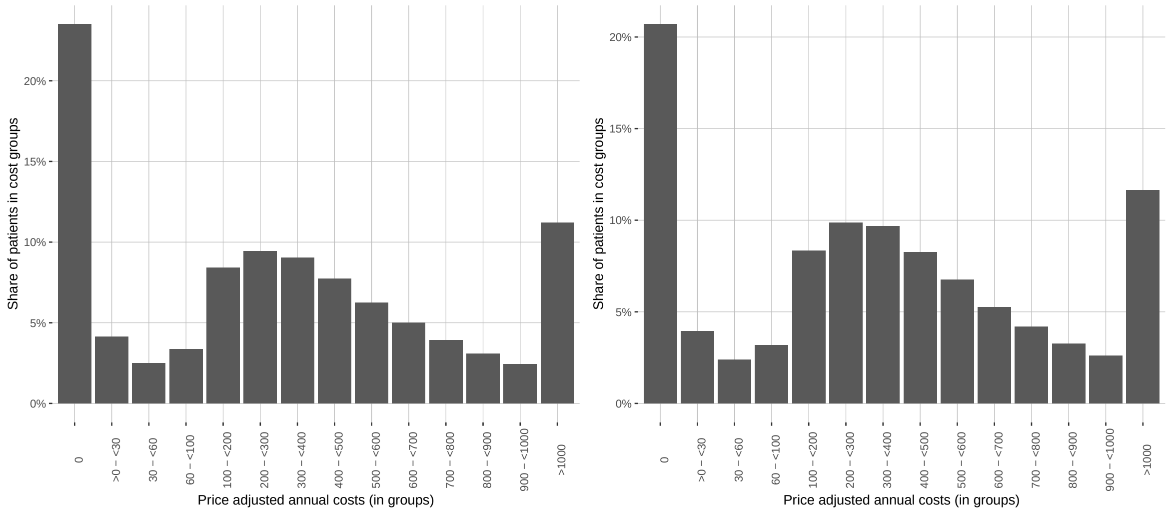
$$y_{it} = \alpha + \beta_0 Cohort_i + \beta_1 Month_i + \beta_2 Cohort_i \times Month_i + Z_{it}\delta + \varepsilon_{it} \tag{5.1}$$

where $Cohort_i$ indicates whether individual i was born between October 1951 and March 1952. The indicator is zero if individual i was born between October 1950 and March 1951. $Month_i$ is the reform indicator that is one if individual i was born between January and March and zero otherwise. $Cohort_i \times Month_i$ is the interaction between the two indicator variables and turns one for every woman born from January 1952 on. Thus, the interaction term marks the individuals who are affected by the reform. In addition, we account for age effects captured in Z_{it}.

The distribution of health costs for the pre-reform cohort (born 1951) at age 59 and 60 (Figure 5.1) shows a strong non-linear pattern. While 20–25% of patients produce zero costs per year, about 50% of patients produce between 100 and 600 Euros costs annually. Due to the non-linearity in the aggregated costs variable and the high share of patients with zero cost we estimate in the main analysis two different models and analyze the effects on the extensive margin and the intensive margin. We estimate the extensive margin in a linear probability model (LPM) in which the outcome variable y_{it} indicates if patients produce costs greater than zero in a given year. For the intensive margin we focus only on positive values and define the outcome y_{it} as the logarithm of the total cost. We also estimate the effect of the overall costs including both the intensive and the extensive margin using the linear costs as an outcome variable. When estimating the effect of the reform on specialist-specific costs, we only focus on the linear model and combine the intensive and extensive margin.[16]

[16] Estimates of the intensive margin would not be comparable across specialists as the share with positive values strongly varies.

Figure 5.1: Cost distributions at age 59 and 60 (birth cohort 1951)



Notes: The left figure presents the costs distribution of women aged 59 years born in 1951. The right figure presents the costs distribution of women aged 60 years born in 1951. Costs are fee-adjusted. *Source:* KBV, own calculations.

In order to identify a causal effects in a difference-and-difference estimator the standard assumptions need to hold. First, the intervention needs to be unrelated to the outcomes at baseline. Since treatment and control group are determined by birthday this assumption is not problematic in our setting. For the same reason the composition of treatment and control group is stable and there are no spillover effects. Secondly, we provide graphical evidence that the parallel trends assumption holds (parallel trends in the outcomes of treatment and control group prior to the intervention) in the Appendix 5.A.3.

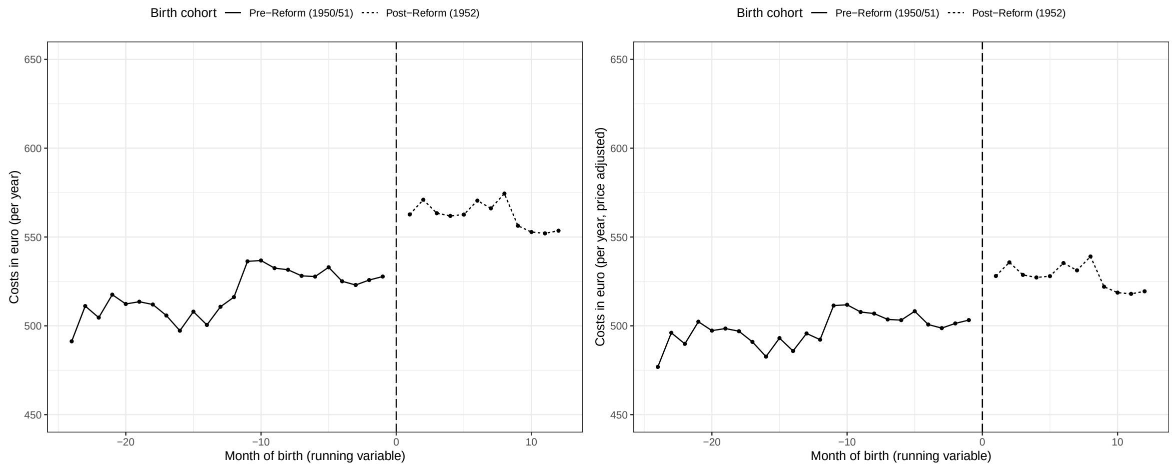
5.5 Results

Before we turn to the discussion of the estimation results, we present graphical evidence on the effect of the pension reform on healthcare costs. Figure 5.2 shows the average healthcare costs per year for women aged 60 to 62 for each birth month.[17] In the left panel we show the raw data. The right hand side presents the adjusted[18] healthcare costs (in fees of year 2009). The vertical lines represent the cutoff date January 1, 1952. For the interpretation, it is important to account for fee changes, since in every year relevant parameters of the healthcare system are adjusted (see Appendix 5.A.1).

[17] In Appendix 5.A.2 Figure 5.A.1, we show the same figures for women in the age range 59-65.

[18] Fees are adjusted to the year 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 (Kassenärztliche Bundesvereinigung KBV, 2019)). For more information on annual changes in the healthcare system, see Appendix section 5.A.1.

Figure 5.2: Annual healthcare costs with and without fee adjustment (1950-52)



Notes: The left figure presents the average healthcare costs per year of women between age 60 and 62 for each birth month. The right figure presents the fee-adjusted average healthcare costs per year (in 2009 fees) of women between age 60 and 62 for each birth month. The vertical lines represent the cutoff date (01/1952).
Source: KBV, own calculations.

Therefore, in the fee-adjusted healthcare costs, the jump between years is smaller. Still, we observe variation in the costs between the months of birth which are related to the seasonality pattern of health (e.g. Boland et al., 2015; Doblhammer and Vaupel, 2001). In the regression analysis, we account for the fee variation and seasonality by using adjusted healthcare costs and using the DiD framework.

Importantly, at the cut-off, we observe the largest jump in healthcare costs: the fee-adjusted costs increase by about 25 euro per person after the cut-off date which corresponds to a relative increase of about 5%. This is first evidence that the increase in the retirement age leads to a sizable increase in healthcare costs. In the following, we turn to the estimation results of the DiD specification to empirically assess this reform effect.

We estimate the effect of the pension reform on healthcare costs on the intensive and extensive margin for different age groups. In Table 5.1 we focus on the intensive margin. In the first Column, we present the results for all women aged 59 – 65. In Column 2, we focus only on women aged 59. Women in this age group were not directly affected by the reform, since retirement via the pension for women was not possible before the age of 60. However, Barschkett et al. (2022) document sizable expectation effects of the 1999 pension reform for several health outcomes, which might affect heath care costs. In Column 3, we consider women aged 60 – 62. Finally, in Column 4 the results for women aged 63 – 65 are presented. These results can be interpreted as

Table 5.1: DiD: Price adjusted annual costs (in logs)

	Dependent variable: Annual costs (in logs)			
	Age: 59-65	Age: 59	Age: 60-62	Age: 63-65
$Cohort_i \times Month_i$	0.022**	0.033**	0.029***	0.011
	(0.008)	(0.010)	(0.008)	(0.009)
$Cohort_i$	0.0003	−0.009	−0.002	0.006*
	(0.004)	(0.009)	(0.004)	(0.003)
$Month_i$	0.016*	0.040***	0.006	0.018*
	(0.007)	(0.007)	(0.007)	(0.008)
Pre-treatment mean	5.884	5.818	5.866	5.926
Age group included	59-65 years	59 years	60-62 years	63-65 years
Control for age	yes	no	yes	yes
Observations	3,294,970	482,177	1,425,656	1,387,137

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Column (1) shows the DiD estimates for women aged 59–65 years, column (2) for women aged 59 years, column (3) for women aged 60–62 years and column (4) for women aged 63–65 years. All specifications include age as control variable, except for column (2). All regressions include the cohort indicator, the reform indicator and their interaction term. Costs are fee-adjusted and in logs (zeros are excluded). *Source:* KBV, own calculations.

post-employment effects since women from the treatment and the control group both have the option to retire.

The estimation results confirm the graphical evidence: We find that the pension reform, i.e., the shift in the retirement age from 60 to 63, increases healthcare costs (Table 5.1). In all specifications (except for 63–65 year old women), the interaction effect that measures the causal effect of the reform, is positive and significant at the 1% or 0.1% level. Specifically, for women aged 59-65 (Column 1), the estimates suggest that the annual healthcare costs increase on average by about 2.2%. According to the linear specification (Table 5.A.1 in the Appendix 5.A.2) this corresponds to an increase of about 14 Euros per person. Note, this effect is smaller than suggested by the graphical evidence, which is due to the seasonality pattern that we account for in the DiD estimation.[19] The effect size over the different age groups is similar. The sizable effect for women aged 59 of over 3% underlines the importance of the expectation effect. At the same time, the insignificant effect on healthcare costs of women aged 63–65 implies that the pension reform did not lead to persistent increases in healthcare

[19]The positive and significant coefficients of the "Month" indicator are in line with the seasonality pattern found by Barschkett et al. (2022). It suggests that women born in the first quarter of the year produce higher healthcare costs than women born in the last quarter of the year.

Table 5.2: DiD: Extensive margin

	Dependent variable: Annual costs (dummy)			
	Age: 59-65	Age: 59	Age: 60-62	Age: 63-65
$Cohort_i \times Month_i$	0.009^{+}	0.019^{**}	0.008	0.006^{+}
	(0.005)	(0.006)	(0.006)	(0.004)
$Cohort_i$	0.020^{***}	0.011^{***}	0.022^{***}	0.021^{***}
	(0.0004)	(0.0004)	(0.0003)	(0.001)
$Month_i$	0.013^{**}	0.003	0.014^{**}	0.016^{***}
	(0.004)	(0.005)	(0.005)	(0.003)
Pre-treatment mean	0.831	0.758	0.808	0.887
Percentage increase in %	1.142	2.558	1.007	0.729
Age group included	59-65 years	59 years	60-62 years	63-65 years
Control for age	yes	no	yes	yes
Observations	3,907,590	627,168	1,737,602	1,542,820

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Column (1) shows the DiD estimates for women aged 59–65 years, column (2) for women aged 59 years, column (3) for women aged 60–62 years and column (4) for women aged 63–65 years. All specifications include age as control variable, except for column (2). All regressions include the cohort indicator, the reform indicator and their interaction term. The outcome variable is a dummy turning 1 if costs are greater than 0 and zero otherwise.
Source: KBV, own calculations.

costs. However, the linear specification (Appendix 5.A.2) suggests a significant increase of about 11 Euros for this age group.

In Table 5.2 we turn to the extensive margin. The results suggest that the increase in healthcare costs can be mostly attributed to increases at the intensive margin. In other words, the additional costs are mainly produced by the group of women with positive costs in absence of the reform. Apart from the age group of 59 year old women, we do not find evidence that the reform induced women to switch from zero healthcare costs to non-zero healthcare costs.

In Figure 5.A.2 we extend the analysis and account directly for the non-linear cost structure documented in Figure 5.1. Specifically, we re-estimate the model with 100 different indicator variables for which we increase the threshold in ten euro increments and present the reform coefficients and confidence intervals. The first coefficient is identical to the extensive margin. Overall, the coefficients have a similar magnitude over the cost distribution but at higher costs the point estimates tend to be smaller but in general they are still significant.

We provide empirical evidence for our identification strategy in Appendix 5.A.3. First, the pre-reform time trends for the treatment and the control groups for the aggregated healthcare costs are very similar (Figure 5.A.3) and, second, the estimates of a placebo test are not significant (Table 5.A.4 for the log-specification and Table 5.A.5 for the extensive margin). Specifically, for the placebo test we use the same empirical specification but artificially shift the design by one year and assign the cohort born in the first quarter 1951 as the treatment group after the hypothetical reform.

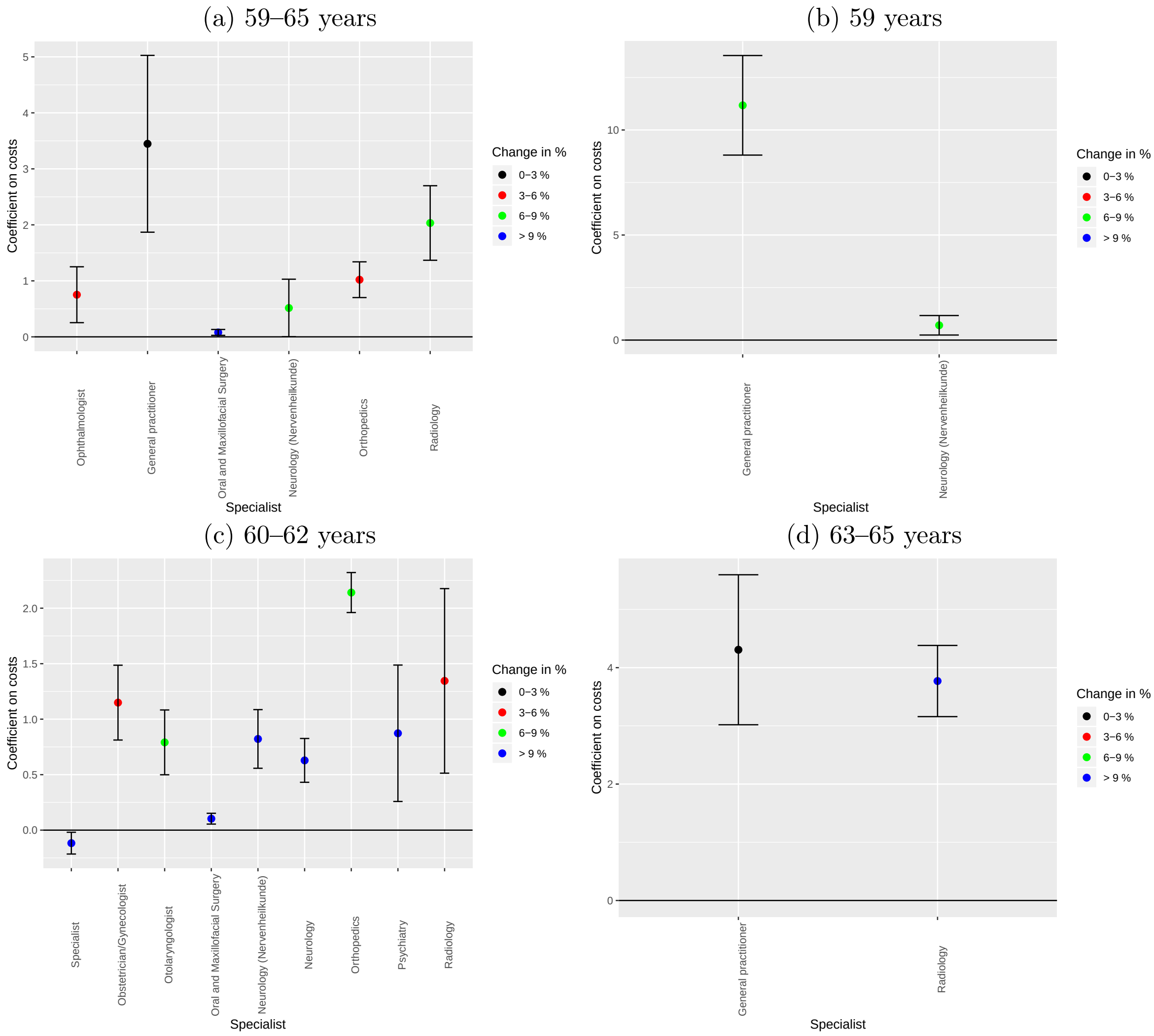
5.5.1 Results by medical specialist

In a next step, we analyze to which specialist care utilization the overall increase in costs can be attributed. This analysis provides insights into whether the increased prevalence of certain diagnoses goes along with increased utilization of the relevant specialists. We present the results for healthcare costs for the 28 different medical specialists that are classified in the data. To correct for multiple hypothesis testing, we adjust the standard errors using a Bonferroni-correction. Figure 5.3 shows the point estimates and 95% confidence bands for the specialists for whom we find significant effects. Costs are fee-adjusted with the same fee index as the overall costs. The complete results for all specialists can be found in the Appendix 5.A.2 in Table 5.A.2 (general fee adjustment) and Table 5.A.3 (adjusted for specialist-specific fees).

Panel (a) shows the results for 59 – 65 year old women. The increase in the retirement age leads to a significant increase in costs for six specialist groups: Ophthalmologists, GPs, oral and maxillofacial surgery, neurology, orthopedics and radiology. The absolute effect is largest for GPs (about 3.5 euro) and thereby contributes about 25% to the increase in the overall costs. In terms of relative price increases, the effects are largest for oral and maxillofacial surgery, neurology, and radiology.

Panel (b) depicts the results for women at age 59 and the bottom left (Panel (c)) and right panel (Panel (d)) for women aged 60 – 62 years and 63 – 65 years, respectively. The results for the 59 year olds suggest, that the increase in the overall costs is mainly driven by increases in the utilization of GP and neurology care. For women aged 60 – 62 years, we find significant increases in the costs for eight specialists: Obstetricians/gynecologists, otolaryngologist, oral and maxillofacial surgery, neurology, orthopedics, psychiatry, and radiology. In absolute terms, the effects are largest for orthopedics (2.1 euro), radiology (1.3 euro) and obstetricians/gynecologists (1.2 euro). Relatively speaking, the rise in costs for oral and maxillofacial surgery, neurology and psychiatry are largest. The costs for specialists decreases due to the reform by 0.1 euro. Similarly to the 59 year old, the increase in overall costs for the 63 – 65 year old seems to be driven by increased utilization of mainly two specialist groups: GPs and radiology.

Figure 5.3: Significant DiD results by medical specialist



Notes: There is a new ("Neurologie") and an old ("Nervenheilkunde") term for the specialist "Neurology". Figures show the statistically significant coefficients (with 95% confidence interval) of the DiD regressions on the specialist specific costs. Standard errors are Bonferroni corrected for multiple hypothesis testing. Panel (a) includes estimates for women aged 59–65 years, panel (b) for women aged 59 years, panel (c) for women aged 60–62 years and panel (d) for women aged 63–65 years.
Source: KBV, own calculations.

Overall, and across the different age groups the results show a relatively clear pattern. We find the strongest increase in the costs for neurology, psychiatry, radiology, GPs, orthopedics and oral and maxillofacial surgery. Based on the data it is not possible to directly identify the mechanism why the costs in the different categories increase. However, the evidence about the effects of the pension reform on health outcomes (Barschkett et al., 2022) allows us to draw indirect conclusions about the mechanisms. Barschkett et al. (2022) document a significant increase in mental health, musculoskeletal diseases, and obesity. Moreover, they find an increase in the number of doctor visits. The increase in mental health can explain the cost effects in neurology and psychiatry which can be related to more frequent doctor visits, diagnostics and treatments. The increase in the costs for orthopedics and radiology are consistent with the finding that musculoskeletal diseases increase. Obesity is often related to mental health and has direct effects on musculoskeletal diseases. Therefore, the increase in obesity is likely to be a driver of the costs effects in the discussed categories. It is difficult to explain the costs effects in oral and maxillofacial surgery based on the mentioned diagnoses. One potential reason for the positive effect of a longer working life on the costs in oral and maxillofacial surgery is that employers may cover part of expensive surgery. The cost effects for the GP can be explained since patients often consult the GP before the specialist.

5.5.2 Costs and revenues of the pension reform

In this section we put our findings into perspective and compare the additional healthcare costs to the overall fiscal effects of the 1999 pension reform. As shown by Geyer and Welteke (2021) the pension reform had a strong negative effect on retirement and a large positive effect on employment as well as on unemployment and inactivity. Specifically retirement rates of affected women decreased by about 25 percentage points. Inactivity and unemployment increased by about ten percentage points, employment by more than 14 percentage points. Geyer et al. (2020) estimate the related short-term effects on government revenues and expenditures which include changes in income taxation, transfer payments and in the social security system. Focusing only on the 1952 cohort and ages 60 to 62, the net effect of the reform amounts to four billion euro.

Relative to this sizable net effect, the additional aggregated healthcare costs are modest. As documented in Table 5.A.1 we find an average increase in health expenditures for women aged 60 to 62 of about 16 euro per year.[20] The cohort size of women born 1952 is about 540,000. Applying the average cost effect and assuming that about 90% of women are covered by the public healthcare system (see section 5.2), the overall healthcare costs related to the pension reform amount to about 7.7 million euro per year in the short run. Thus, relative to the fiscal net effect of four billion euro, the healthcare costs amount to less than 2%. This cost effect is a lower bound as our data only covers outpatient care. Yet, since the pension reform mainly affected mental

[20]This estimate needs to be interpreted as an intent to treat effect (ITT) since not all women were eligible for the pension for women. According to Geyer and Welteke (2021) about 60% of women in the cohorts considered were affected by the pension reform reform.

health, musculoskeletal diseases, and obesity (Barschkett et al., 2022) healthcare costs related to outpatient care are - at least in the short run - of central importance.

5.6 Conclusion

In this paper, we document that an increase in the retirement age leads to a significant increase in healthcare costs. To identify the causal effect of the increase in the retirement age, we exploit a cohort-specific pension reform which increased the early retirement age by three years between women born in two adjacent cohorts. The analysis is based on data that include the universe of individuals insured through the German public healthcare system (almost 90% of the German population) and comprises a ten-year observation period (2009 – 2018). Our results show that healthcare costs increase overall by about 2.9% for women in the age group directly affected by the increase in the retirement age (60-62). Moreover, we find expectation effects for women at the age of 59 and indirect post-employment effects for women between 63 and 65. In addition, we show that the cost increase is mainly driven by increased utilization of the following specialist groups: ophthalmologists, general practitioners (GPs), neurology, orthopedics and radiology. The absolute effect is largest for GPs (about 3.5 euro) and thereby contributes about 25% to the increase in the overall costs.

While the effects are significant and meaningful on the individual level, we show that the increase in healthcare costs are modest relative to the positive fiscal effects of the pension reform. Specifically, we estimate an aggregate increase in the outpatient costs of about 7.7 million for women aged 60-62 and born in 1952. Relative to the corresponding estimate of the net effects of the pension reform of about four billion euro (Geyer et al., 2020) this translates into a relative effect of less than 2%.

Thus, from an aggregate perspective, our results of an increase in the healthcare costs do not provide strong evidence against an increase in the retirement age. However, the increase of costs on the individual level support the findings of previous studies focusing on individual health outcomes, that positive fiscal effects of a longer working life can be counteracted by potential negative health consequences for individuals. Our cost estimation focuses on the public healthcare costs and abstracts from individual disutility or other disadvantages due to worse health as well as other societal costs such as a decrease in labor productivity or an increase in sickness absence at work. For political decisions on retirement ages, such non-monetary factors also need to be taken into account.

5.A Appendix

5.A.1 Prices and quantities in the German healthcare system

Every medical service that is covered by public health insurance is valued by a point system (Einheitlicher Bewertungsmaßstab – EBM). Every year, the National Association of Statutory Health Insurance Physicians (KBV) and the National Association of Statutory Health Insurance Funds (GKV Spitzenverband) negotiate at the federal level about the money value of the valuation point (price component) and the morbidity trends (quantity component). Following federal negotiations, the respective regional associations negotiate the specific terms for each region, such as e.g., the regional prices and morbidity parameters that determine the total compensation package.[21]

The total compensation package for outpatient services in each region is financed by the health insurance providers. The respective total compensation packages are split into two parts: the morbidity-related compensation package (MGV) and the extra-budgetary compensation package (EGV).

Medical service providers in the public healthcare system settle their quarterly accounts with their regional Association of Statutory Health Insurance Physicians (KV) based on the point system[22] and regional prices. Around 70 percent[23] of the medical services are paid from MGV. Since funds in MGV are fixed and limited, medical service providers get paid less than the negotiated rate if they exceed their quarterly ceiling.[24] Specific services (such as e.g., certain vaccinations) are always covered according to EBM and paid from the EGV budget.

Within the legal framework, the Federal Joint Committee (Gemeinsamer Bundesausschuss; G-BA) decides on questions of coverage of the public health insurance in Germany. This board consists of representatives of public health insurance providers and medical service providers (BMG, 2020).

[21] The information in this section is collected from GKV (2021); Kassenärztliche Bundesvereinigung KBV (2021a,b,c).

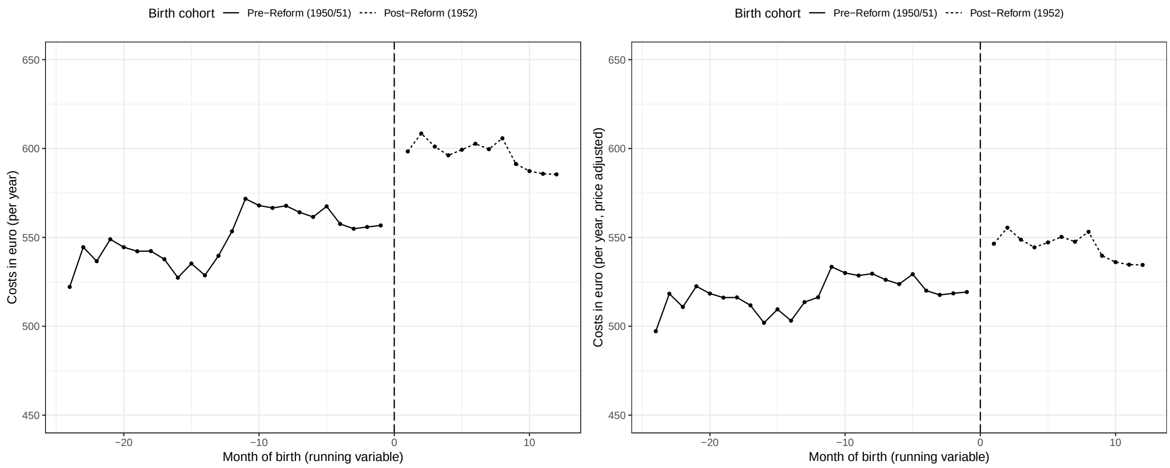
[22] For more details, see BMG (2016).

[23] See, e.g., GKV (2021).

[24] Since 2012, the specific rules for the distribution of MGV funds to medical service providers are set by the regional KVs; see Neumann et al. (2014); Walendzik and Wasem (2019).

5.A.2 Additional results

Figure 5.A.1: Annual healthcare costs with and without fee adjustment (1950-52)



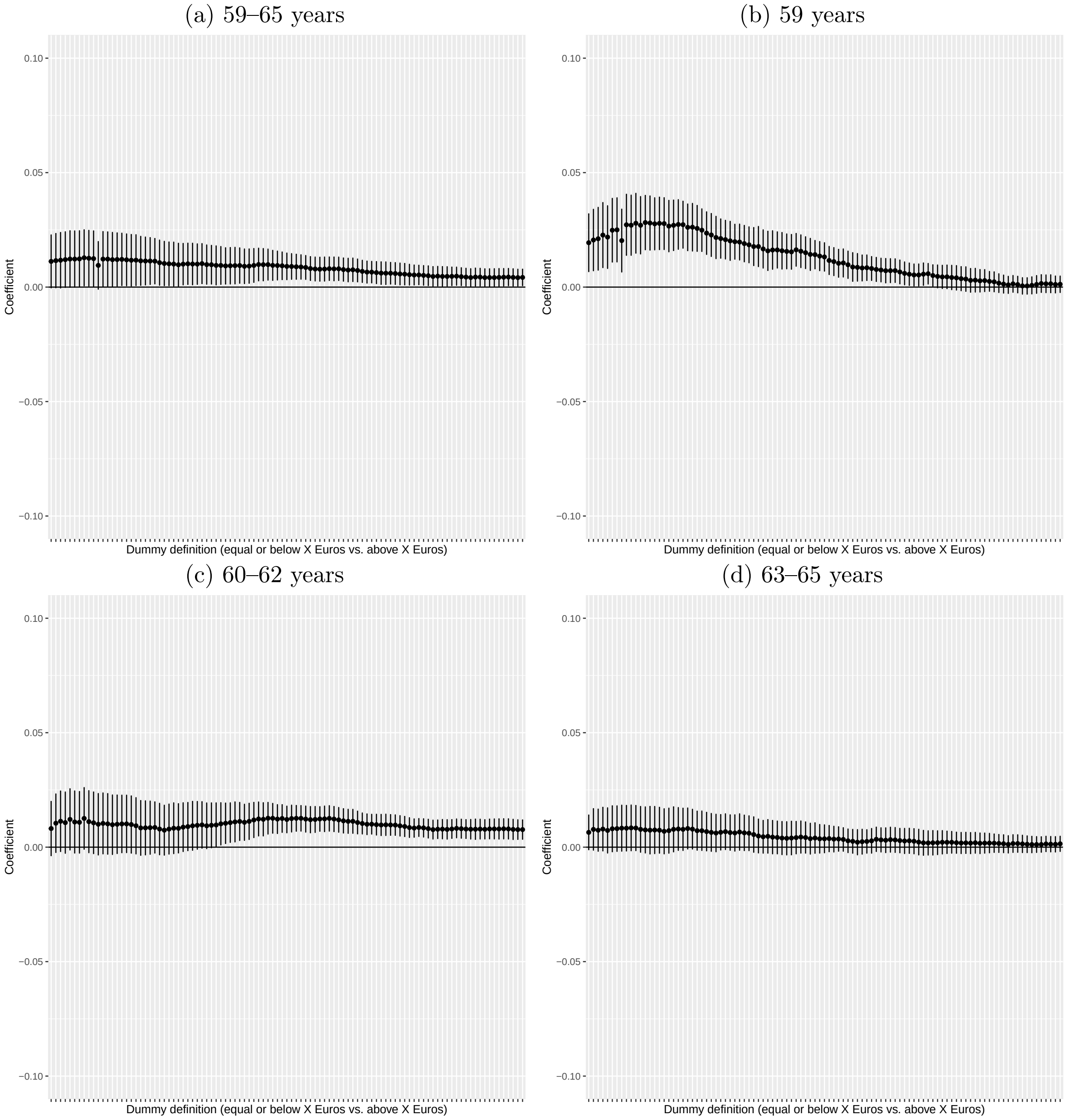
Notes: The left figure presents the average healthcare costs per year of women between age 59 and 65 for each birth month. The right figure presents the fee adjusted average healthcare costs per year (in 2009 fees) of women between age 59 and 65 for each birth month. The vertical lines represent the cutoff date (01/1952).
Source: KBV, own calculations.

Table 5.A.1: Linear specification: Price adjusted annual costs

	Dependent variable: Annual costs			
	Age: 59-65	Age: 59	Age: 60-62	Age: 63-65
$Cohort_i \times Month_i$	14.08*	13.87*	16.28*	11.27*
	(5.63)	(5.77)	(6.57)	(5.36)
$Cohort_i$	9.66***	3.65	9.48***	12.33***
	(2.58)	(3.29)	(2.73)	(2.93)
$Month_i$	17.55***	20.73***	13.34*	21.35***
	(5.14)	(5.16)	(6.14)	(4.33)
Pre-treatment mean	517.86	459.30	498.49	563.57
Percentage increase in %	2.72	3.02	3.27	2.00
Age group included	59-65 years	59 years	60-62 years	63-65 years
Control for age	yes	no	yes	yes
Observations	3,907,590	627,168	1,737,602	1,542,820

Notes: +p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Column (1) shows the DiD estimates for women aged 59–65 years, column (2) for women aged 59 years, column (3) for women aged 60–62 years and column (4) for women aged 63–65 years) All specifications include age as control variable, except for column (2). All regressions include the cohort indicator, the reform indicator and their interaction term. Costs are fee-adjusted.
Source: KBV, own calculations.

Figure 5.A.2: Effects along the cost distribution



Notes: Coefficients from estimating 100 times Equation 5.1 with different definitions of the outcome variable. In the first regression, the outcome variable is defined as an indicator variable, taking value zero if healthcare costs are zero and one if healthcare costs are greater than zero, i.e. the extensive margin. In the second regression, the indicator is one if costs $\leq$ 10 euros and one if costs $>$ 10 euros. In the third regression, the indicator is one if costs $\leq$ 20 euros and one if costs $>$ 20 euros. The threshold increases with increments of 10 euros up until 1000 euros.
Source: KBV, own calculations.

Table 5.A.2: DiD: Annual costs by medical specialist (general fee-adjustment)

	Dependent variable: Annual costs			
	Age:59-65	Age:59	Age:60-62	Age:63-65
Anesthesiology	0.20	−0.27	0.46**	0.10
	(0.19)	(0.22)	(0.17)	(0.25)
MHT adjusted p-value	1	1	0.155	1
Pre-treatment mean	7.17	6.75	7.01	7.52
Change in %	2.82	-4.03	6.55	1.39
Dermatologist	0.06	0.30+	0.17	−0.16
	(0.15)	(0.18)	(0.13)	(0.19)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	10.69	9.09	10.05	12.05
Change in %	0.56	3.35	1.68	-1.34
General practitioner	3.45***	11.17***	−0.10	4.31***
	(0.70)	(1.21)	(0.69)	(0.63)
MHT adjusted p-value	0.000	0.000	1	0.000
Pre-treatment mean	165.01	142.61	161.39	178.17
Change in %	2.09	7.84	-0.06	2.42
Human genetics	−0.26*	−0.28	−0.25	−0.26
	(0.12)	(0.19)	(0.19)	(0.23)
MHT adjusted p-value	0.923	1	1	1
Pre-treatment mean	1.18	0.83	0.94	1.60
Change in %	-21.84	-33.19	-27.02	-15.98
Internal Medicine	2.77	3.63	2.10	3.18
	(2.45)	(2.50)	(3.20)	(2.71)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	66.36	55.18	63.01	74.68
Change in %	4.18	6.58	3.33	4.26
Laboratory	0.58**	0.80**	0.65*	0.42
	(0.20)	(0.27)	(0.27)	(0.28)
MHT adjusted p-value	0.113	0.082	0.397	1
Pre-treatment mean	16.51	13.30	15.29	19.18
Change in %	3.54	6.00	4.27	2.19
Medical psychotherapist	0.61	0.12	1.01*	0.35
	(0.39)	(0.46)	(0.43)	(0.42)
MHT adjusted p-value	1	1	0.531	1

Table 5.A.2: DiD: Annual costs by medical specialist (general fee-adjustment)

	Dependent variable: Annual costs			
	Age:59-65	Age:59	Age:60-62	Age:63-65
Pre-treatment mean	4.85	5.45	5.12	4.29
Change in %	12.53	2.19	19.78	8.11
Neurology (Nervenheilkunde)	0.52**	0.70***	0.82***	0.09
	(0.17)	(0.18)	(0.13)	(0.26)
MHT adjusted p-value	0.049	0.003	0.000	1
Pre-treatment mean	8.54	8.12	8.47	8.79
Change in %	6.04	8.68	9.70	1.04
Neurology	0.30*	0.00	0.63***	0.06
	(0.13)	(0.16)	(0.10)	(0.21)
MHT adjusted p-value	0.498	1	0.000	1
Pre-treatment mean	5.04	4.25	4.50	5.97
Change in %	6	0.04	13.95	0.97
Non-medical psychotherapist	0.54	0.85	0.98	−0.09
	(0.60)	(0.83)	(0.69)	(0.65)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	13.44	16.24	14.11	11.54
Change in %	3.99	5.22	6.95	-0.81
Nuclear medicine	0.13	−0.14	0.25	0.11
	(0.09)	(0.12)	(0.16)	(0.10)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	7.00	5.68	6.57	7.90
Change in %	1.88	-2.41	3.74	1.39
Obstetricians/Gynecologists	0.52**	−0.24	1.15***	0.11
	(0.19)	(0.29)	(0.16)	(0.27)
MHT adjusted p-value	0.189	1	0.000	1
Pre-treatment mean	28.75	29.18	28.22	29.16
Change in %	1.80	-0.84	4.07	0.39
Ophthalmology	0.75***	−0.00	1.00**	0.78*
	(0.16)	(0.23)	(0.34)	(0.37)
MHT adjusted p-value	0.000	1	0.098	0.988
Pre-treatment mean	21.89	13.68	18.54	29.00
Change in %	3.44	-0.02	5.40	2.68
Oral and Maxillofacial Surgery	0.08***	−0.06	0.10***	0.11**
	(0.02)	(0.08)	(0.02)	(0.04)

Table 5.A.2: DiD: Annual costs by medical specialist (general fee-adjustment)

	Dependent variable: Annual costs			
	Age:59-65	Age:59	Age:60-62	Age:63-65
MHT adjusted p-value	0.003	1	0.000	0.195
Pre-treatment mean	0.63	0.49	0.58	0.75
Change in %	12.51	-12.04	17.92	14.41
Orthopedics	1.02***	−0.59+	2.14***	0.41
	(0.15)	(0.36)	(0.10)	(0.26)
MHT adjusted p-value	0.000	1	0.000	1
Pre-treatment mean	29.32	27.47	29.08	30.34
Change in %	3.48	-2.13	7.36	1.36
Other physicians	0.13	−1.72	−0.77+	1.90*
	(0.20)	(1.14)	(0.42)	(0.78)
MHT adjusted p-value	0.166	0.032	1	0.38
Pre-treatment mean	14.57	10.11	13.41	17.69
Change in %	0.90	-16.96	-5.77	10.76
Other service providers	−0.65	−0.97	1.37	−2.79+
	(1.44)	(1.25)	(1.71)	(1.50)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	27.25	25.83	26.79	28.35
Change in %	-2.37	-3.74	5.10	-9.84
Otolaryngologist	0.38**	0.07	0.79***	0.03
	(0.14)	(0.16)	(0.14)	(0.20)
MHT adjusted p-value	0.233	1	0.000	1
Pre-treatment mean	10.32	8.90	9.72	11.58
Change in %	3.64	0.75	8.14	0.29
Pathology	−0.00	−0.20**	−0.01	0.08
	(0.04)	(0.07)	(0.09)	(0.05)
MHT adjusted p-value	1	0.062	1	1
Pre-treatment mean	5.37	4.63	5.19	5.87
Change in %	-0.05	-4.29	-0.10	1.37
Phoniatrics Pedaudiology	0.00	−0.00	−0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	0.19	0.09	0.14	0.27
Change in %	0.67	-3.04	-4.47	4.18

Table 5.A.2: DiD: Annual costs by medical specialist (general fee-adjustment)

	Dependent variable: Annual costs			
	Age:59-65	Age:59	Age:60-62	Age:63-65
Physical rehabilitation medicine	0.10^{+}	0.13	0.16^{*}	0.03
	(0.06)	(0.10)	(0.08)	(0.08)
MHT adjusted p-value	1	1	0.861	1
Pre-treatment mean	2.01	1.76	1.93	2.21
Change in %	5.16	7.19	8.48	1.24
Psychiatry	0.41^{**}	-0.06	0.87^{***}	0.08
	(0.15)	(0.21)	(0.24)	(0.10)
MHT adjusted p-value	0.199	1	0.005	1
Pre-treatment mean	4.67	5.58	4.75	4.22
Change in %	8.77	-1.09	18.37	1.90
Radiology	2.03^{***}	-0.32	1.35^{***}	3.77^{***}
	(0.32)	(0.37)	(0.34)	(0.32)
MHT adjusted p-value	0.000	1	0.002	0.000
Pre-treatment mean	28.07	25.40	27.44	29.85
Change in %	7.24	-1.26	4.90	12.63
Radiotherapy	0.13	-1.72	-0.77^{+}	1.90^{*}
	(0.20)	(1.14)	(0.42)	(0.78)
MHT adjusted p-value	1	1	1	0.38
Pre-treatment mean	14.57	10.11	13.41	17.69
Change in %	0.90	-16.96	-5.77	10.76
Specialists	-0.07^{*}	-0.03	-0.12^{***}	-0.04
	(0.03)	(0.03)	(0.04)	(0.03)
MHT adjusted p-value	0.297	1	0.019	1
Pre-treatment mean	0.68	0.76	0.75	0.57
Change in %	-10.67	-3.38	-15.70	-7.16
Surgery	-0.04	-0.31	0.31	-0.32
	(0.25)	(0.29)	(0.37)	(0.22)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	16.26	15.07	16.24	16.77
Change in %	-0.23	-2.05	1.94	-1.92
Urology	0.17	0.22^{*}	0.27^{+}	0.04
	(0.17)	(0.10)	(0.16)	(0.23)
MHT adjusted p-value	1	0.043	1	1

Table 5.A.2: DiD: Annual costs by medical specialist (general fee-adjustment)

	Dependent variable: Annual costs			
	Age:59-65	Age:59	Age:60-62	Age:63-65
Pre-treatment mean	4.57	4.01	4.26	5.14
Change in %	3.67	5.38	6.23	0.74
Age group included	59-65 years	59 years	60-62 years	63-65 years
Observations	3,904,369	627,097	1,737,117	1,540,155

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. There is a new ("Neurologie") and an old ("Nervenheilkunde") term for the specialist "Neurology". Standard errors are clustered on month of birth and robust. Column (1) shows the DiD estimates for women aged 59–65 years, column (2) for 59 year old women, column (3) for 60–62 year old women and column (4) for 63–65 year old women. All specifications include age as control variables (except for column (2)). All regressions include the cohort indicator, the reform indicator and their interaction term. Costs are fee-adjusted.
Source: KBV, own calculations.

Table 5.A.3: DiD: Annual costs by medical specialist (specific price adjustment)

	Dependent variable: Annual costs			
	Age: 59-65	Age: 59	Age: 60-62	Age: 63-65
Anesthesiology	0.23	0.00	0.42*	0.10
	(0.20)	(0.23)	(0.18)	(0.28)
MHT adjusted p-value	1	1	0.460	1
Pre-treatment mean	7.52	6.75	7.28	8.10
Change in %	3.02	0.06	5.77	1.24
Dermatologist	0.08	0.25	0.18	−0.11
	(0.14)	(0.18)	(0.13)	(0.18)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	10.40	9.09	9.81	11.59
Change in %	0.75	2.69	1.85	-0.92
General practitioner	3.63***	8.75***	0.60	4.97***
	(0.69)	(1.20)	(0.67)	(0.61)
MHT adjusted p-value	0.000	0.000	1	0.000
Pre-treatment mean	161.44	142.61	158.29	172.62
Change in %	2.25	6.14	0.38	2.88

Table 5.A.3: DiD: Annual costs by medical specialist (specific price adjustment)

	Dependent variable: Annual costs			
	Age: 59-65	Age: 59	Age: 60-62	Age: 63-65
Human genetics	-0.29^{*}	-0.30	-0.32	-0.25
	(0.13)	(0.20)	(0.21)	(0.24)
MHT adjusted p-value	0.660	1	1	1
Pre-treatment mean	1.24	0.83	1.06	1.62
Change in %	-23.23	-36.37	-30.00	-15.47
Internal Medicine	2.94	4.32^{+}	2.71	2.63
	(2.68)	(2.58)	(3.43)	(3.11)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	71.74	55.18	65.83	85.11
Change in %	4.09	7.83	4.11	3.09
Laboratory	0.64^{**}	0.45^{+}	0.85^{***}	0.49^{+}
	(0.20)	(0.27)	(0.26)	(0.28)
MHT adjusted p-value	0.035	1	0.025	1
Pre-treatment mean	16.15	13.30	14.58	19.06
Change in %	3.97	3.41	5.80	2.55
Medical psychotherapist	0.67	0.26	1.05^{*}	0.40
	(0.43)	(0.46)	(0.47)	(0.47)
MHT adjusted p-value	1	1	0.707	1
Pre-treatment mean	5.19	5.45	5.44	4.81
Change in %	12.84	4.80	19.36	8.23
Neurology (Nervenheilkunde)	0.53^{***}	0.69^{***}	0.83^{***}	0.12
	(0.15)	(0.18)	(0.12)	(0.24)
MHT adjusted p-value	0.014	0.002	0.000	1
Pre-treatment mean	8.03	8.12	8.00	8.06
Change in %	6.59	8.47	10.45	1.50
Neurology	0.26^{*}	-0.32^{+}	0.63^{***}	0.09
	(0.12)	(0.17)	(0.09)	(0.20)
MHT adjusted p-value	0.891	1	0.000	1
Pre-treatment mean	4.95	4.25	4.59	5.65
Change in %	5.30	-7.52	13.63	1.60
Non-medical psychotherapist	0.53	1.10	0.99	-0.22
	(0.65)	(0.87)	(0.74)	(0.72)
MHT adjusted p-value	1	1	1	1

Table 5.A.3: DiD: Annual costs by medical specialist (specific price adjustment)

	Dependent variable: Annual costs			
	Age: 59-65	Age: 59	Age: 60-62	Age: 63-65
Pre-treatment mean	14.33	16.24	14.97	12.85
Change in %	3.72	6.79	6.63	-1.68
Nuclear medicine	0.12	-0.23^{+}	0.26	0.09
	(0.09)	(0.12)	(0.17)	(0.11)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	7.36	5.68	6.96	8.50
Change in %	1.58	-4.00	3.80	1.05
Obstetricians/Gynecologists	0.42^{*}	-0.06	0.75^{***}	0.24
	(0.20)	(0.29)	(0.17)	(0.27)
MHT adjusted p-value	0.862	1	0.000	1
Pre-treatment mean	29.39	29.18	29.22	29.66
Change in %	1.43	-0.22	2.57	0.82
Ophthalmology	0.69^{***}	0.21	0.74^{*}	0.81^{*}
	(0.16)	(0.23)	(0.34)	(0.34)
MHT adjusted p-value	0.000	1	0.780	0.439
Pre-treatment mean	21.17	13.68	18.45	27.27
Change in %	3.24	1.56	4.01	2.98
Oral and Maxillofacial Surgery	0.07^{***}	-0.02	0.09^{***}	0.10^{**}
	(0.02)	(0.07)	(0.02)	(0.03)
MHT adjusted p-value	0.001	1	0.000	0.137
Pre-treatment mean	0.57	0.49	0.54	0.65
Change in %	12.79	-4.36	16.50	14.67
Orthopedics	1.01^{***}	-0.82^{*}	2.20^{***}	0.41
	(0.17)	(0.36)	(0.10)	(0.29)
MHT adjusted p-value	0.000	0.642	0.000	1
Pre-treatment mean	31.19	27.47	30.77	33.16
Change in %	3.23	-3.00	7.14	1.24
Other physicians	0.23^{*}	0.63^{**}	0.60^{**}	-0.35^{**}
	(0.10)	(0.20)	(0.19)	(0.12)
MHT adjusted p-value	0.616	0.056	0.041	0.071
Pre-treatment mean	7.64	7.42	7.52	7.87
Change in %	3.00	8.43	7.97	-4.46
Other service providers	-1.37	-4.25^{**}	0.68	-2.49^{*}

Table 5.A.3: DiD: Annual costs by medical specialist (specific price adjustment)

	Dependent variable: Annual costs			
	Age: 59-65	Age: 59	Age: 60-62	Age: 63-65
	(1.34)	(1.45)	(1.65)	(1.27)
MHT adjusted p-value	1	0.089	1	1
Pre-treatment mean	26.67	25.83	29.66	23.64
Change in %	-5.12	-16.45	2.30	-10.55
Otolaryngologist	0.36*	−0.04	0.74***	0.10
	(0.15)	(0.16)	(0.14)	(0.21)
MHT adjusted p-value	0.360	1	0.000	1
Pre-treatment mean	10.48	8.90	9.96	11.70
Change in %	3.44	-0.47	7.43	0.83
Pathology	0.02	−0.01	0.04	0.01
	(0.04)	(0.06)	(0.08)	(0.05)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	5.22	4.63	5.01	5.70
Change in %	0.40	-0.28	0.78	0.25
Phoniatrics Pedaudiology	0.00	0.00	−0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	0.19	0.09	0.14	0.27
Change in %	1.37	0.54	-4.60	4.86
Physical rehabilitation medicine	0.10	0.01	0.20*	0.02
	(0.06)	(0.11)	(0.09)	(0.09)
MHT adjusted p-value	1	1	0.526	1
Pre-treatment mean	2.22	1.76	2.11	2.54
Change in %	4.48	0.41	9.56	0.88
Psychiatry	0.43**	−0.06	0.92***	0.09
	(0.15)	(0.21)	(0.23)	(0.10)
MHT adjusted p-value	0.119	1	0.002	1
Pre-treatment mean	4.66	5.58	4.70	4.24
Change in %	9.32	-1.09	19.47	2.21
Radiology	2.03***	−0.32	1.35***	3.77***
	(0.32)	(0.37)	(0.34)	(0.32)
MHT adjusted p-value	0.000	0.134	0.000	0.000
Pre-treatment mean	28.07	25.40	27.44	29.85

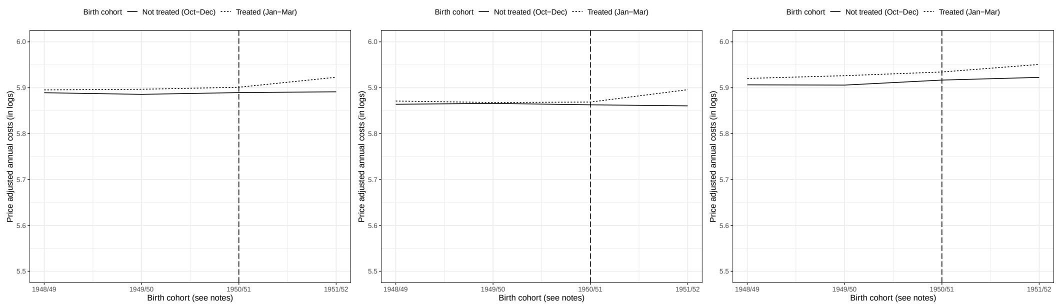
Table 5.A.3: DiD: Annual costs by medical specialist (specific price adjustment)

	Dependent variable: Annual costs			
	Age: 59-65	Age: 59	Age: 60-62	Age: 63-65
Change in %	7.24	-1.26	4.90	12.63
Radiotherapy	0.24	-0.85	-0.90^{*}	1.97^{*}
	(0.21)	(1.12)	(0.43)	(0.81)
MHT adjusted p-value	1	1	0.945	0.385
Pre-treatment mean	14.88	10.11	13.50	18.37
Change in %	1.62	-8.43	-6.67	10.75
Specialists	-0.07^{*}	-0.03	-0.12^{***}	-0.04
	(0.03)	(0.03)	(0.04)	(0.03)
MHT adjusted p-value	0.297	1	0.019	1
Pre-treatment mean	0.68	0.76	0.75	0.57
Change in %	-10.67	-3.38	-15.70	-7.16
Surgery	-0.06	-0.13	0.20	-0.33
	(0.26)	(0.30)	(0.38)	(0.24)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	0.15	0.09	0.27	0.03
	(0.19)	(0.10)	(0.18)	(0.26)
Urology	0.17	0.22^{*}	0.27^{+}	0.04
	(0.17)	(0.10)	(0.16)	(0.23)
MHT adjusted p-value	1	1	1	1
Pre-treatment mean	4.99	4.01	4.63	5.80
Change in %	2.92	2.18	5.89	0.46
Age group included	59-65 years	59 years	60-62 years	63-65 years
Observations	3,904,369	627,097	1,737,117	1,540,155

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. There is a new ("Neurologie") and an old ("Nervenheilkunde") term for the specialist "Neurology". Standard errors are clustered on month of birth and robust. Column (1) shows the DiD estimates for women aged 59–65 years, column (2) for 59 year old women, column (3) for 60–62 year old women and column (4) for 63–65 year old women. All specifications include age as control variables (except for column (2)). All regressions include the cohort indicator, the reform indicator and their interaction term. Costs are specialist-specific fee-adjusted.
Source: KBV, own calculations.

5.A.3 Robustness

Figure 5.A.3: DiD Graphs for 60–65, 60–62 and 63–65 year old women



Notes: The left figure presents the average annual costs of women between age 60 and 65 for each birth cohort, the middle figure the costs for women aged 60 to 62 and the right figure for women aged 63 to 65 years. The vertical lines represent the cutoff date (01/1952). Birth cohort 1948/49 represents women born between October to December 1948 (control group) and January and March 1949 (treatment group). Accordingly, birth cohorts 1949/50 represent women born between October to December 1949 and January and March 1950, birth cohorts 1949/50 represent women born between October to December 1950 and January and March 1951 and birth cohorts 1951/52 represent women born between October to December 1951 and January and March 1952.
Source: KBV, own calculations.

Table 5.A.4: Placebo-DiD: Price adjusted annual costs (in logs)

	Dependent variable: Annual costs		
	Age: 60-65	Age: 60-62	Age: 63-65
$Cohort_i \times Month_i$	0.001	0.004	-0.003
	(0.009)	(0.010)	(0.010)
Pre-treatment mean	5.893	5.866	5.921
Age group included	60-65 years	60-62 years	63-65 years
Control for age	yes	yes	yes
Observations	2,797,163	1,416,699	1,380,464

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Column (1) shows the Placebo-DiD estimates for women aged 60–65 years, column (2) for women aged 60–62 years and column (3) for women aged 63–65 years). In the Placebo specification the reform date was artificially shifted to one year earlier (01/1951). All specifications include age as control variable. All regressions include the cohort indicator, the reform indicator and their interaction term. Costs are fee-adjusted and in logs. Zero costs are excluded.
Source: KBV, own calculations.

Table 5.A.5: Placebo-DiD: Extensive margin

	Dependent variable: Annual costs		
	Age: 60-65	Age: 60-62	Age: 63-65
$Cohort_i \times Month_i$	0.001	0.005	−0.003
	(0.005)	(0.007)	(0.003)
Pre-treatment mean	0.894	0.844	0.952
Age group included	60-65 years	60-62 years	63-65 years
Control for age	yes	yes	yes
Observations	3,129,333	1,678,753	1,450,580

Notes: $^{+}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered on month of birth (running variable) and robust. Column (1) shows the Placebo-DiD estimates for women aged 60–65 years, column (2) for women aged 60–62 years and column (3) for women aged 63–65 years). In the Placebo specification the reform date was artificially shifted to one year earlier (01/1951). All specifications include age as control variable. All regressions include the cohort indicator, the reform indicator and their interaction term. The outcome variable is a dummy turning 1 if costs are greater than 0 and zero otherwise.
Source: KBV, own calculations.

CHAPTER 6

Costs and short-term effects of a home-visiting program in *BRISE* – first steps for a cost-effectiveness analysis[1]

6.1 Introduction

It is broadly established that socio-economic inequalities emerge during early childhood or even during prenatal phases (e.g., Currie and Almond, 2011; Fernald et al., 2013; Silvestrin et al., 2020). We also know that investments during the early stages of a child's life can effectively mitigate these inequalities (e.g., Heckman et al., 2010). This knowledge is also reflected in a sharp rise in public spending on early childhood programs in most OECD countries (OECD, 2021). For example, in Germany, public spending almost doubled from 19.5 billion in 2010 to 36.9 billion Euros in 2020 (Destatis, 2020)[2] – a trend that will likely continue in the future. Regarding the effectiveness of specific policies, there is a large body of research focusing on early childcare programs aiming to nurture child development directly (e.g., Barnett, 1985; Havnes and Mogstad, 2011b; Heckman et al., 2013; Karoly et al., 2006; Kautz et al., 2014). However, substantially less research exists on the effects of interventions that target parenting skills and knowledge as a way to improve child development and parental well-being (e.g., Camehl et al., 2020; Klebanov et al., 2001; Olds, 2006).[3] The most prominent type of

[1]This chapter is joint work with Laura Schmiz (DIW Berlin) and Sophia Schmitz (BiB Wiesbaden). We are grateful to the research associates at the University of Bremen for data access and for their excellent support. We further thank C. Katharina Spiess, Andrew Judy, Louisanne Knierim, Jonas Jessen, as well as the participants at internal seminars at DIW Berlin, BRISE consortium meetings, the 2021 meeting of the scientific advisory board of BRISE and the 2022 GEBF conference. Moreover, we gratefully acknowledge funding from the BRISE project (project number: FKZ: 01NV1601A-G).

[2]These figures correspond to the total expenditure on educational institutions under public and private sponsorship for children under six years.

[3]See Heckman and Mosso (2014) for an overview of the existing studies.

these programs are home-visiting schemes which advise parents - usually mothers - on aspects of everyday life with an infant, such as mother-and-child interaction, nutrition, and ways to seek support when needed (e.g., Camehl et al., 2020). Another aspect that is often inadequately addressed by previous studies is the efficiency of such programs, i.e., how the costs relate to the benefits. In the face of scarce public resources, efficiency studies are an important tool for policymakers and can help making investments in early childhood programs more compelling (e.g., Karoly, 2012; Spieß, 2013). Studies considering the costs and effectiveness of interventions are particularly scarce in the German and European context, leading to a lack of evidence-based accountability for early childhood investments.

This paper presents novel evidence on the short-term effectiveness and costs of the home-visiting program *Pro Kind* which targets parenting styles and parental behavior during pregnancy and early childhood. It entails bi-weekly home visits starting prenatally and lasting until children turn two. The main interest of this study is to understand whether *Pro Kind* already had a significant impact on mother and child outcomes during the first seven months of the children's lives. *Pro Kind* is the first program within a systematic chain of home- and center-based preschool interventions established to support disadvantaged families from pregnancy until school entry under the Bremen Initiative to Foster Early Childhood Development ("Bremen Initiative zur Stärkung frühkindlicher Entwicklung" – *BRISE*).[4]

To establish causality, we exploit the fact that treatment status of BRISE was randomly assigned on the neighborhood level. The treatment for the families recruited in treatment areas (N=124) comprises better access and specific information on the home-based programs, while participation barriers for recruited families in the control neighborhoods (N=176) were much higher. In our empirical analysis we apply a combination of methods: We exploit the random treatment assignment in an instrumental variables (IV) approach to overcome endogeneity in the participation decision. To overcome any remaining endogeneity concerns, we combine the IV approach with entropy balancing methods to account for the observed differences in socio-economic characteristics between participating families from treatment and control neighborhoods. To analyze costs, we use comprehensive, self-collected data from a yearly cost survey following the ingredient method (Levin and McEwan, 2000) to set up a micro cost data set

[4]Launched in 2017 in a large Western German city (Bremen), the idea of *BRISE* is to systematically combine existing home- and center-based preschool programs into an intervention chain to support disadvantaged families from pregnancy until school entry. Thus, *BRISE* examines whether integrating separate schemes leads to higher cumulative effects on child development while also being less expensive. Two other programs of BRISE starting before age three, *Opstapje* and *Tipp Tapp* are not evaluated, since the data to analyze *Opstapje* is not available yet and *Tipp Tapp* participation was not randomized (BRISE Consortium, 2022; Schütte et al., 2020).

covering all costs related to the program implementation. We then proceed to compare the overall cost per child per year to other well-established early childhood programs that entail parenting skills elements.

BRISE collects data on an extensive range of outcomes covering different aspects related to maternal and child well-being, providing a comprehensive picture of the early impact of the program. For *mothers*, we focus on a number of behavioral outcomes – smoking, alcohol consumption, breastfeeding, and soothing strategies – as well as maternal well-being, including postnatal depression, measured by the Edinburgh Postnatal Depression Scale (EPDS) (Cox et al., 1996), and perception of change in living conditions. Both nicotine and alcohol consumption during pregnancy and breastfeeding is associated with several adverse infant health outcomes (e.g., Polanska et al., 2015). Similarly, maternal well-being is a strong predictor of child development (e.g., Berger and Spiess, 2011; Dahlen, 2016), positive health outcomes (e.g., Diener and Chan, 2011) and labor productivity (e.g., Oswald et al., 2015). Regarding *child* outcomes, we look at the Milestones of Normal Development in Early Years (MONDEY) score, a comprehensive child development indicator developed by psychologists (Pauen, 2011; Pauen et al., 2012).

Our results concerning the very short-term impact of *Pro Kind* on child and maternal outcomes display statistically insignificant effects. Due to the small sample sizes and many missing observations, the effects are imprecisely estimated, which does not allow us to draw conclusions. Put aside the data issues, it is not surprising to find no significant impact on child and mother outcomes in the child's first months of life. It may take several years for the effects of early childhood programs to manifest. For example, studies that evaluated *Pro Kind* in other contexts find minor improvements in some maternal and child outcomes, which materialize earliest at 12 months (e.g., Jungmann et al., 2015; Sandner, 2013, 2019; Sandner et al., 2018). Hence, future studies evaluating the effectiveness of *Pro Kind* at later stages are in a better position to conduct a sound analysis. Our cost analysis reveals that the average costs per participant per year in *Pro Kind* range between 3,468 and 3,861 Euros over the study period (2017–2020). About 80% of the total costs can be attributed to personnel costs. In comparison to other programs, *Pro Kind* belongs to the less costly programs – only one program displays lower costs per child per year (Sure Start in the UK, Cattan et al., 2021), while the prominent Perry Preschool program (Heckman et al., 2010) is about 5.5 times as expensive as *Pro Kind* .

Our study makes several contributions. First, our study adds to the literature on the effects of early childhood education and care programs (e.g., Barnett, 2011; Cunha et al., 2010; Heckman and Mosso, 2014), especially those targeting parenting skills (e.g., Camehl et al., 2020; Lindsay et al., 2011). More specifically, we add to the growing evidence on the effectiveness of home-visiting programs (e.g., Doyle et al., 2015; Olds, 2006; Sandner, 2019) – as opposed to center-based programs (e.g. Triple P, Camehl et al., 2020) – on improving child and maternal outcomes. So far, studies of this kind have primarily focused on Anglo-American programs (e.g., Heckman et al., 2017; Love et al., 2005; Olds, 2006), with little evidence existing in the European and German context. Exceptions include evaluations of the Swiss counterpart of Parents as Teachers (PAT) (Schaub et al., 2019), the German family-supporting prevention program "Keiner fällt durchs Netz" (KfdN; "Nobody Slips Through the Net") (Sidor et al., 2013) and previous research on *Pro Kind* in three German federal states (e.g., Jungmann et al., 2015; Sandner, 2013, 2019; Sandner et al., 2018). While the program has been evaluated with respect to its effectiveness on several maternal and child outcomes, this is the first study to estimate effects on the MONDEY score (e.g., Pauen, 2011; Pauen et al., 2012) and the Edinburgh Postnatal Depression Scale (EPDS) (Cox et al., 1996). The MONDEY score is an observational tool to document early childhood development until age three, comprising eight domains, e.g., gross and fine motor skills and visual perception. As such, it is an extremely comprehensive indicator of early childhood development. The EPDS is a widely established and reliable measure of postnatal depression. In total, we consider an extensive range of outcomes covering different aspects related to maternal and child well-being, providing a comprehensive picture of the early impact of the program. Lastly, we make use of a unique cost panel data set based on a cost survey developed by family and education economists (Barschkett and Schmitz, 2020). The cost database breaks down costs into different components and records a 100% response rate of the program providers on a yearly basis. As such, this study builds a valuable basis for future even more detailed cost-efficiency and cost-benefit studies within BRISE.

The remainder of this paper is structured as follows. Section 6.2 summarizes the related literature. Next, we describe the institutional background, including the *BRISE* programs. Section 6.4 presents our data set, control and outcome variables. After that, we outline our empirical strategy and provide descriptive statistics. Section 6.6 presents our empirical findings on the effectiveness of the programs and costs. Lastly, Section 6.7 concludes.

6.2 Literature

There is a large and growing literature focusing on the effectiveness of early childcare programs (e.g., Baker et al., 2008; Carneiro and Ginja, 2014; Cornelissen et al., 2018; Currie and Almond, 2011; Heckman et al., 2010). Especially targeted programs such as the Perry Preschool Program and the Abecedarian Project, including both center- and home-based elements, were found to have large medium and long-term effects on participants' (non-)cognitive and labor market outcomes (e.g., Barnett and Masse, 2007; Campbell and Ramey, 1991; Heckman et al., 2010). However, also universal programs, such as Sure Start, a center-based early childhood intervention in the UK offering a range of services including childcare and parenting support, were shown to positively affect children's health (Cattan et al., 2021) and other related outcomes such as family functioning and behavioral problems (e.g., Sammons et al., 2015). Similarly, Triple P (Positive Parenting Program), a multi-level parenting and family support strategy mostly relying on center-based interventions, had positive effects on child and maternal outcomes (e.g., Camehl et al., 2020; Kim et al., 2018). Slightly less empirical research exists on the effectiveness of home-based programs that seek to improve child development through enhanced parenting skills. These programs are referred to as home-visiting, prevention, or home-based parenting programs. They usually target low-SES families during pregnancy or shortly after childbirth and consist of regular home visits by nurses, social workers, or paraprofessionals. Table 6.1 offers a summary of selected home-visiting programs in terms of their effects on child and maternal outcomes. The studies are chosen based on program similarities to *Pro Kind* (home-based, age group) and the way effects are measured (age at measurement, outcomes, (quasi-)experimental design).[5]

A large majority of these home-visiting programs that have been evaluated in terms of their effectiveness are located in the United States. Famous examples include the Nurse Family Partnership (NFP) (Olds et al., 2002) and Early Head Start (Love et al., 2005). Many of these programs have proven to be effective concerning child development, maternal outcomes, and parenting behavior (Table 6.1), although most effects could only be measured in the medium to long term. For example, Early Head Start parents showed higher emotional engagement and support (e.g., Love et al., 2005) and the NFP led to a reduction in smoking, increased maternal employment, and more mother-child interactions (Olds et al., 2002) and increased home investments, parenting attitudes and mental health for mothers of infants at age two (Heckman, 2007). Similar results

[5] We do not claim to offer a complete overview. Other (less recent) literature overviews can be found in, for example, Sweet and Appelbaum (2004), Peacock et al. (2013) and MacMillan et al. (2009).

were found for the programs Minding the Baby (Sadler et al., 2013) and Parents as Teachers (PAT) (Wagner and Clayton, 1999). One notable exception where significant effects on parenting behavior were measured at a very early stage is the Home visiting program in Queensland: At six weeks, mothers receiving the home visits had significant reductions in postnatal depression, demonstrated more positive interactions with their infants, and achieved higher scores in maternal-infant secure attachment, among others (Armstrong et al., 1999).

Table 6.1: Overview of the effects of selected home-based early childhood interventions

	Child development			Maternal outcomes					
Program	Cognitive	Socio-emotional	Health	Mental health	Health behavior	Employment	Fertility	Parenting	Age at measurement
Anglo-American programs									
Nurse Family Partnership	+○	+○		+	+	+	-		0-2, 6, 12 y.
Preparing for Life (PFL)			+						6, 12, 18, 24, 36 m.
Early Head Start	+	+	○	+					3 y.
Minding the Baby			○	○			-	+	4, 12, 24 m.
CCDP	○	○	○	○	○	○		○	3-5 y.
Home visiting (Queensland)			○	+				+	6 w.
Home visiting (UK)			○	○				+	2, 6, 12 m.
FDRP	+	+				○			3 y.
PAT (USA)			○					+	2-3 y.
European programs									
PAT (Switzerland)	+○	+	○						3 y.
KfdN		+						+	1 y.
Pro Kind									
- All	+○	○	○	+	+	-	+		0-2 y.
- Girls	+								0-2 y.
- Boys	○								0-2 y.

Note: The table reports effects of home-visiting programs found in the literature, where + indicates positive effects (improvement), ○ indicates null effects, - indicates negative effects (deterioration). The following papers found the basis of this depictions: NFP: Olds et al. (2002), Heckman et al. (2017), Olds (2006); PFL: Doyle et al. (2015); Early Head Start: Love et al. (2005); Minding the Baby: Sadler et al. (2013); CCDP: St. Pierre and Layzer (1999); Home visiting (Queensland): Armstrong et al. (1999); Home visiting (UK): DuMont et al. (2010), McIntosh et al. (2009); FDRP: Besharov et al. (2011); PAT (USA): Wagner and Clayton (1999); PAT (Switzerland): Schaub et al. (2019); KfdN: Sidor et al. (2013); Pro Kind: Sandner et al. (2018), Sandner (2019), Sandner (2013), Sandner and Jungmann (2017), Jungmann et al. (2015).
Source: Own depiction.

While positive effects were also recorded on child outcomes, they usually take even longer to materialize. For example, at three years old, Early Head Start participants demonstrated improved cognitive and language development (Love et al., 2005). Regarding the NFP in Memphis, participants showed enhanced cognitive skills for both genders and improved socio-emotional skills for females at age six (Heckman, 2007). At the age of three, children in primarily Spanish-speaking Latino communities showed significant gains in cognitive, communication, social, and self-help development as a result of participating in PAT (Wagner and Clayton, 1999). The Irish program Preparing for Life (PFL) (Doyle et al., 2015) recorded positive effects on child health after 18 and 24 months. At six months, a small effect was found for the level of immunizations. Another large-scale home-visiting program, the Comprehensive Child Development Program (CCDP), did not record any statistically significant effects on either child or mother outcomes (St. Pierre and Layzer, 1999).

Less evidence is available on the effectiveness of home-visiting programs outside the Anglo-American region. The Swiss counterpart to the PAT showed improved children's adaptive behavior, developmental status, and language skills at the age of three (Schaub et al., 2019). The German family-supporting prevention program "Keiner fällt durchs Netz" (KfdN; "Nobody Slips Through the Net") targets psychosocially stressed families in a controlled trial setting. Sidor et al. (2013) shows that children in the intervention group showed improved social development scores and were judged by their mothers to be less "difficult." In addition, the dysfunctionality of the mother-child interaction was reduced in this group compared to the control group. No intervention effects were found for the degree of maternal stress or maternal sensitivity towards the child.

Multiple studies have evaluated the effectiveness of *Pro Kind* in improving different parental and child outcomes at different points in time. For example, Jungmann et al. (2015) find minor positive effects on parental self-efficacy, feelings of attachment, social support, and maternal oral health. The effects were measured at different times (between pregnancy and age two of the children). However, up until age two, the authors do not find economically meaningful effects on maternal or child health. Sandner (2013) find small increases in infants' cognitive developments at 12 months, which fade out at 24 months. For other outcomes, they do not find effects at this early stage (Sandner, 2013). Sandner and Jungmann (2017) add that the effects on cognitive development were driven by girls, explained by greater parental investment for girls than for boys. Concerning health, Sandner et al. (2018) find positive effects on maternal mental health but no effect on other health outcomes such as healthcare utilization, health behaviors, and physical health of mother and child (measured at multiple points in time between pregnancy and age two of the child). At 36 months prenatally, Sandner (2019) points out a decrease in maternal employment, an increase in subsequent birth, and positive effects on maternal well-being and life satisfaction. These findings suggest that the *Pro Kind* program entails mid or long-term benefits rather than short-term effects.

Home-visiting programs are labor-intensive and hence produce high costs. Therefore, cost-effectiveness or cost-benefit studies are essential for policymakers to justify the introduction or continuation of such programs. However, evidence on both costs and effectiveness of home-visiting programs is scarce. So far, cost-effectiveness studies focusing on home-visiting schemes during early childhood mainly exist in the Anglo-

American context (Schmitz and Kröger, 2017).[6] The most widely researched program in this respect is the Nurse Family Partnership (NFP) program in the US that has been implemented at multiple sites (e.g., Glazner et al., 2004; Karoly, 2017; Karoly et al., 1998; Miller and Hendrie, 2015). The analyses are net-cost analyses from the standpoint of government spending, i.e., the benefits to the government are monetized in the way that the government needs to provide less support to nurse-visited families (e.g., reduction in social and health care costs). In the long run, NFP's benefits exceed the costs; however, this threshold is only reached in children's teenage years (Glazner et al., 2004; Karoly, 2017; Karoly et al., 1998). Additionally, a home-visiting program in the UK was shown to improve maternal sensitivity and infant cooperativeness at an incremental cost of 3264 pounds per woman (Barlow et al., 2007; McIntosh et al., 2009).

We contribute to this small but growing strand of the literature by focusing on the German context. We analyze both the effectiveness as well as incurring costs for *Pro Kind*. In Europe, such programs' (cost-)effectiveness is likely different from the US-American case. In most European countries, an established social welfare system exists in addition to such programs. Furthermore, in the US, childcare as well as parenting support programs often target the most in-need groups, whereas European countries usually adopt a universal childcare approach. We examine the effectiveness and costs of a targeted parenting scheme in an institutional context with a generous social welfare system and universal childcare.

6.3 Institutional Background

6.3.1 Design of BRISE

Launched in 2017, the project *BRISE* is a cooperation of the city of Bremen with several research institutions.[7] The idea of the *BRISE* program is to systematically combine existing home-based and preschool programs into a chain of intervention measures to support families from pregnancy until school entry (see Figure 6.A.1). Findings in the international literature suggest that time-limited individual measures often have only

[6] One of the reasons for this is that North American countries have traditionally spent less on their welfare systems than European countries, creating the need for evidence on the efficiency of alternative policies (Korpi and Palme, 1998).

[7] The scientific consortium is composed of the University of Bremen, Leibniz Institute for Science and Mathematics Education (IPN), the German Institute for Economic Research (DIW), The Federal Institute for Population Research (BiB), the Leibniz Institute for Research and Information in Education (DIPF), the Max-Planck Institute for Education Research, the Leibniz Institute for Educational Trajectories (LIfBi), University of Heidelberg, and the University of Bamberg.

small effects (e.g., Puma et al., 2012) and that periods of lack of support during the first years of a child's life should be avoided (e.g., Schütte et al., 2020). As a result, the systematic integration of regionally already established programs into an intervention chain is a potentially promising and cost-effective means to achieve sustainable positive effects on child development. BRISE tests this hypothesis in a large-scale quasi-experimental field study conducted in close collaboration with the political and administrative level in a city whose population structure is characterized by a substantial proportion of socially and culturally disadvantaged families. Measured against the national median, Bremen has the highest at-risk-of-poverty rate of all German states (IAW, 2018). On average, Bremen records one of the highest shares of students who fall below minimum requirements both in primary (Stanat et al., 2017) and middle school (Stanat et al., 2019). Within the scope of *BRISE,* the city has expanded the provision of programs belonging to the support chain. Specifically, 70 additional places have been created for *Pro Kind*, corresponding to an expansion of 78 percent (Schütte et al., 2020).

BRISE was rolled out in 27 districts in Bremen, which were selected based on data from the State Statistical Office to identify districts with a comparatively high proportion of disadvantaged families. Specifically, the following indicators were taken into account: the proportion of the population under the age of fifteen receiving basic unemployment benefits (*Arbeitslosengeld II*), the proportion of the population with a qualifying school-leaving certificate, and the proportion of the population with an immigrant background. Furthermore, the political relevance was considered; on the one hand, the rate of relocations from the district was included here, and on the other hand, the birth rate. In addition, the implementation conditions were taken into account, such as the existing infrastructure of programs for early childhood development and whether there is political interest and willingness to reorganize the local supply structure. A large number of local actors familiar with the situation on site were actively involved in this process. A network of multipliers has also been established to recruit participating families (Schütte et al., 2020). The 27 BRISE districts were then randomly divided into treatment (10) and control (17) districts such that the two groups were similar in terms of the criteria mentioned above. As shown in Table 6.A.1, at the time this selection was made, there was no statistically significant difference between the two groups. One exception is the number of births, which was slightly higher in the control districts. *BRISE* targets low-SES expectant parents, which are ideally recruited during the last trimester of pregnancy but at the latest when the children are ten weeks old. Families were recruited based on the following eligibility criteria:

one or both parent(s) should have a migration background[8], or low education[9], or low income. Another precondition was that the children should not be severely ill upon program entry, i.e., during the end of pregnancy or shortly after postpartum. Since the program had difficulties recruiting enough families, however, these criteria became less strict over time. In total, *BRISE* has recruited 404 families with 456 participating children, following a randomized controlled trial (RCT) design with randomization at the neighborhood level.

When comparing the *BRISE* sample with a representative sample of the German population (SOEP)[10] (Table 6.A.2), we see that *BRISE* families are on average slightly more disadvantaged than SOEP families. Specifically, there is a mixed picture regarding the educational status of the mothers: While *BRISE* mothers are more likely to have no school degree, there is also a higher share of mothers with an academic school degree (Abitur) in the *BRISE* groups than in the average SOEP respondent mother. One potential explanation for the relatively high share of highly educated mothers could be that we have a high share of mothers with a foreign school degree which complicates the comparison of education levels. In most remaining SES characteristics, the *BRISE* sample appears indeed more disadvantaged than the average SOEP respondent. For example, *BRISE* families have a lower net income and more often have an immigrant background.

The treatment can be regarded as a combination of access and information treatment: *BRISE* provides financial support to the treatment districts to scale up their offer of the programs forming the intervention chain. In addition, family counselors (researchers of the University of Bremen) inform the families in the treatment neighborhoods about the relevant *BRISE* programs, arrange contacts to the practice sites, and support the families of the treatment group in taking advantage of the continuous support. Figure 6.A.2 gives an example of a flyer handed out to the families. BRISE children in the treatment districts are guaranteed a place in the programs. In sum, the process is designed to keep the practical costs and administrative barriers for families to register as low as possible. The families in the control group are free to use these programs but receive no such information or organizational support. Both groups have access to other regular German healthcare services or other programs.

[8]This criteria was defined as fulfilled if at least one parent was born outside of Germany.

[9]Low education was defined as having less than a high-school degree, i.e., no school degree or a lower secondary school degree.

[10]We select a sample that includes families in urban areas, with children below one, and excludes the SOEP migration (M1-M5), low-income (L2) samples and all observations surveyed before 2010.

Three home visiting schemes are offered during the first year of *BRISE*, forming the first part of the intervention chain. The following section describes these programs, focusing on *Pro Kind*, which is the subject of this study.

6.3.2 Pro Kind (PK)

Pro Kind, run by the German Red Cross, is modeled after the US Nurse-Family-Partnership (NFP) project, which has been shown in evaluations to be effective in improving child and maternal outcomes (for an overview, see Eckenrode et al., 2017). Like NFP, *Pro Kind* begins in the last trimester of pregnancy and ends when the child turns two. During pregnancy, first time mothers are counseled about nutrition, the importance of avoiding alcohol and nicotine, and typical warning signs of complications. After childbirth, *Pro Kind* staff, i.e., midwives, pediatric nurses, or social pedagogues, provide counseling about appropriate childcare and soothing techniques, health, and nutrition, as well as how to interact with the child in a way that promotes child development. The frequency of the home visits varies between weekly, biweekly, and monthly, summing up to around 52 home visits, each lasting on average 90 minutes. The *Pro Kind* teaching materials are closely linked to NFP guidelines and structure the theme of each home visit, although *Pro Kind* staff may decide to adapt the contents to the specific needs of the families (Sandner et al., 2018). In addition, *Pro Kind* covers public transportation costs to prenatal checkups and hands out monetary thank-you gifts of 25 Euros for participation in the interviews.

6.3.3 Other home visiting programs in *BRISE*

Opstapje is also administered by the German Red Cross and is a home-based intervention for socioeconomically disadvantaged families, focusing on improving parent-child interactions by strengthening parenting skills and resources in the home (Sann and Thrum, 2005). *Opstapje Baby* usually begins when the child is two months old and ends when they turn three years. In Bremen, however, *Opstapje* is designed to start when the child turns six months (BRISE Consortium, 2013). Since we currently only have data up to the age of seven months, we cannot evaluate *Opstapje* in terms of its effectiveness yet.

Tipp Tapp takes up the postnatal concept of NFP and offers early prevention in at-risk families. Here, parents receive counseling from a nurse at three points in time (after birth, after six months, and after one year) as part of an announced home visit. The counseling covers nutritional issues, care, design of the child's living environment, accident prevention, prophylaxis, vaccinations, and participation in early detection programs. Unfortunately, a causal evaluation of *Tipp Tapp* is impossible since the program proactively contacts expectant low-SES families in all of Bremen, not distinguishing between treatment and control neighborhoods. As a result, the setting does not meet the standards for a quasi-experimental evaluation design. Consequently, participation in *Tipp Tapp* only enters our analysis as a control variable (e.g., BRISE Consortium, 2013).

6.4 Data

The sample used in this study comprises the first 300 children (born 2017–2019) and their families who participated in *BRISE*.[11] Families were surveyed between 2017 and 2020. The *BRISE* project conducts regular surveys in the participating households that build the basis for this evaluation. After the families apply for the program and are considered a potential fit based on the official criteria, a screening interview takes place surveying essential SES background variables. When the final decision is made, and the family is officially part of the program, they are surveyed again shortly after giving birth (t0), and when the child is three months (t1), seven months (t2), and twelve months (t3) old, respectively.[12] The survey instruments are mainly based on well-established questions from the Germany Socio-Economic Panel (SOEP) (Goebel et al., 2019) and the National Educational Panel Study (NEPS) (Blossfeld and Von Maurice, 2011), combined with specific questions for the *BRISE* project.

6.4.1 Outcome variables

Our outcome variables on maternal and child outcomes are currently only available for the first two waves (children's age: 3 months (t1) and 7 months (t2)). In the analysis, we standardize all outcome variables except for binary variables.[13]

Mother outcomes

We consider six variables describing maternal behavior and well-being. First, *BRISE* surveys smoking and alcohol consumption during pregnancy and after childbirth. Both

[11]Recruitment only ended in 2022. Thus, future studies can draw on the full sample of families participating in BRISE.

[12]More regular surveys occur at later stages until the child turns six. However, this study focuses on the first year after childbirth; hence, only these first five interviews up to t3 are relevant.

[13]In many of our outcome variables, we observe a non-trivial amount of missing values. These appear to be random in terms of socio-economic characteristics but negatively correlate with the participation in *Pro Kind*. Tables showing the correlations between missings in the outcomes, the control variables, treatment status and the program participation are available upon request.

nicotine and alcohol consumption during pregnancy and breastfeeding is associated with several adverse infant health outcomes (e.g., Polanska et al., 2015). In both cases, the outcome variable takes value one when the mother indicates that, at the time of the interview, she regularly smokes or drinks alcohol, respectively. Second, mothers are also asked about breastfeeding habits. Our outcome variable takes value one if the mother indicates that she is still *breastfeeding* at t1 or t2. Third, we evaluate maternal "*soothing strategies*". *BRISE* surveys whether and how often mothers apply each of the following techniques: "carrying the baby," "leaving the baby cry," "giving the baby medication," "smacking the baby," "breastfeeding," "cradling the child in one's arms," "shaking the baby," "playing music," and "singing for the baby." The variables range on a standard Likert scale (Likert, 1932) from value 0 (never) to value 5 (multiple times per day). We recode the variables such that higher values indicate better soothing strategies and build an index by adding the different items. The index ranges between 32 and 54, with a mean of 48.4 in t1 and 47.0 in t2.

Fourth, besides these variables covering different aspects of the maternal behavior towards the child, we also focus on outcomes related to *maternal well-being.* Besides the individual and societal relevance of well-being, e.g., enhancing productivity (DiMaria et al., 2020), maternal well-being is also linked to child development, e.g., improving verbal skills and reducing socio-emotional problems (Berger and Spiess, 2011). An essential aspect of this is whether the mother has experienced postnatal depression. A total of ten items, e.g., "I was feeling so sad that I had trouble sleeping" and "Things became too much for me," yield the Edinburgh Postnatal Depression Scale (Cox et al., 1996). This scale ranges from 0 (no depression) to 30 (very high risk of depression). As suggested by Cox et al. (1996), we build an indicator variable taking the value 1 for values between 0 and 9 (low risk of depression), 2 for values between 10 and 12 (medium risk of depression), and 3 for all values greater or equal to 13 (high risk of depression). Additionally, we analyze whether the mother perceived a change in her living conditions. This category comprises nine items, which we recode such that they are all positively phrased and add up to build an index.[14].

Child outcomes

"MONDEY" (Pauen, 2011; Pauen et al., 2012) includes a description of 111 mile-

[14]Change in living conditions consists of the following items: 1. "Raising my child brings me joy," 2. "I am often at the end of my strength," "I am satisfied with my new role as a mother," 4. "I often do not feel up to the new tasks and requirements," 5. "I am concerned about my child's health," 6. "My living conditions have changed very much," 7. "Giving my child much tenderness is very important to me," 8. "I suffer from being limited to my role as a mother," 9. "I also get to know others through the child and make new contacts ."These items range on a four-point Likert scale from value 0 (do not agree at all) to 3 (completely agree) (Likert, 1932; Siegle, 2020). The resulting index ranges between 17 and 32 has a mean of 26.4 in t1 and 27.1 in t2.

stones, each assigned to one of eight areas to measure child development (i.e., gross and fine motor development, perception and cognition, language, social relations, self-regulation, and emotions). Together these eight areas offer a standardized inventory to monitor child development from zero to three. MONDEY is conceptionally similar to the Bayley Scales of Infant Development (BSID) (Benson and Haith, 2010). As such, MONDEY is a comprehensive measure for child development that is exceptionally suited to track infant development during the first year of the children's lives. Many other child development indicators only focus on later outcomes. We use standardized developmental scores for each of these eight areas and an overall average score as our outcome variables.[15]

6.4.2 Program participation

BRISE collects very detailed data on the *Pro Kind* participation of the families. Specifically, we have information for each family on the number of home visits, the date of the first and the last visit, whether there was an interruption of program participation, and if so, during which period, and the age of the children when they started and finished the program. This information is collected by the providers of *Pro Kind*. Based on this information, we build a treatment variable. In order to do so, first, we define the number of scheduled visits at each survey time (15 at t1 and 23 at t2, assuming that visits start three months before birth and take place bi-weekly). Second, we define the individual maximum number of visits possible at each survey time by taking the age of children at the program start into account. Third, we estimate the individual participation rate by dividing the number of home visits by the individual maximum number of visits. In the main specification, the participation rate serves as our measure for program participation. In a robustness check, we employ a dummy variable that takes value one if the participation rate is at least 50% and zero otherwise.

6.4.3 Control variables

In our empirical analysis, we control for the following individual and household characteristics: Age of the child in days on the day of the survey, sex of the child, a dummy taking the value one if it is the first-born child of the family, a dummy for whether a doctor assessed the pregnancy to be a high-risk pregnancy, household net income[16],

[15]However, since we do not find effects in any of the separate developmental areas, we only present the results on the overall score (Table 6.4).

[16]Due to issues with missing values, a part of the observations of family income are imputed. We use mean values at different educational levels of the mother to impute missing values. We then divide

and parental characteristics[17]. Besides the mother's age, we include the school degree of the mother[18], the training level of the mother[19], the labor market participation of both parents during the screening phase, i.e., pre-birth[20]. In addition, we control for migration background[21] and self-rated satisfaction of the mother with the family situation during the screening interview, i.e., before the treatment started. Our set of control variables includes the most important socio-economic characteristics which are either exogenous or measured before the treatment occured.

Descriptive statistics of outcome and control variables and program participation for the full sample of families are depicted in column 2 of Table 6.2. In the full sample, 8–9% of children participate in *Pro Kind* whereof the majority lives in treatment districts. The relatively low participation rate may be explained by the fact that only first time mothers (56% of the sample) are eligible for *Pro Kind* participation. The means of the outcomes are reported for both measurement times: three months (t1) and seven months (t2). While smoking is relatively constant across measuring times (12–13%), the share of mothers who drink alcohol doubles from t1 to t2. Postnatal depressions, change in living conditions and soothing strategies are relatively constant across measuring times. Except for soothing strategies in t2, there are no statistically significant differences in the outcomes between treatment and control group.

6.5 Empirical strategy

6.5.1 Effectiveness analysis

In a first step, we estimate simple OLS regressions of the following form:

$$y_i = \beta_1 + \beta_2 PK_i + X_i'\beta_3 + \mu_i \tag{6.1}$$

the observed and imputed observations into six income categories: below 750 euros (1), 750-1500 euros (2), 1500-2500 euros (3), 2500-3500 euros (4), 3500-5000 euros (5), and over 5000 euros (6).

[17] The control variables except for the labor force status are only available for the mother.

[18] Here, we distinguish between having no school degree (1) as the base category having a general school degree (2), i.e., below 12 years of education, and having a high school degree (3), i.e., having obtained the German *(Fach-)Abitur* or completed at least 12 years of schooling within or outside of Germany.

[19] In this variable, no training (1) forms the base category, apprenticeship/technical training takes value 2, and academic degrees takes value 3.

[20] Here, we apply the following categories: not being in the labor force (1) as the base category, working part-time (2), and working full-time (3).

[21] This variable takes value zero if no parent has a migration background, one if one parent was either born abroad or has an indirect migration background, and two if the latter applies to both parents.

where y_i are the different child and maternal outcome variables. The variable of interest, program participation (PK_i), is a continuous variable that indicates the share of *Pro Kind* visits child i participated in, ranging from zero to 100 percent.[22] X_i' is our vector of control variables, as described in section 6.4. However, employing the OLS model in Equation 6.1 does not necessarily produce estimates that can be interpreted as causal. Identifying a causal effect of program participation on child and maternal outcomes faces potential endogeneity threats. The choice for participating might be influenced by unobserved characteristics that also affect the outcome variables, causing an omitted variables bias. One example of such an unobserved variable is the parents' openness to advice in parenting style. This degree of openness could influence the likelihood of participation and directly affect our outcomes (e.g., maternal smoking behavior). Another threat could be reverse causality; for example, maternal well-being might influence how much support from such programs mothers need and thus demand. Thus, estimating Equation 6.1 might lead to a biased and inconsistent estimate of program participation and would not reflect a causal effect.[23]

To overcome these endogeneity issues, we exploit the fact that BRISE randomly assigned an information and access treatment on the neighborhood level. Formally, we exploit this random variation within an instrumental variable (IV) framework (e.g., Angrist et al., 1996). Thus, we predict the variation in program participation using the assigned treatment status on the neighborhood level as an instrument that determines the endogenous regressor (PK_i) but only affects the dependent variables (y_i) through its effect on this independent variable (program participation).

Validity of the instrument. In order for the access and information treatment to qualify as a valid instrument, it must fulfill several conditions: The relevance and the exogeneity assumptions. Relevance means that the instrument must sufficiently correlate with the endogenous regressor, i.e., program participation. Arguably, the access and information treatment satisfies the relevance condition as it enhances the popularity of the programs and lowers entry barriers. The correlation between the instrument and the endogenous regressor is empirically tested in the first stage regression, where the endogenous variable is regressed on the instruments and the exogenous covariates

[22] Note, the variable can take values above 100 percent as we assumed conservative number of maximum visits. Thus, families participating very regularly, can achieve more than the assumed maximum number of visits.

[23] There are reasons to expect both upwardly and downwardly biased OLS estimates. For example, if only mothers open for advice participate, we expect the OLS estimator to be upward biased. Alternatively, if we expect mothers with low subjective well-being to be more likely to seek support and thus participate, the OLS estimator would be downward biased. We cannot account for the endogeneity issues by including all confounding factors as control variables, as some of them are not observed in the data at hand or might be unknown.

(columns 1 and 4 in Table 6.4). The robust first stage F-statistics displayed in the main regression table in section 6.6 (Table 6.4) are mostly around 15 to 20. This result supports our argument.

The more critical assumption is the exogeneity assumption of the instrument, which requires that the instrument is not correlated with the error term and thus influences the outcome variable only through the endogenous regressor. It seems plausible that the access and information treatment influences our outcomes only through program participation if the randomization worked perfectly and the treatment and control group are balanced in their socio-economic characteristics. On the neighborhood level, the randomization worked well, i.e., there are no statistically significant differences in socio-economic characteristics between treatment and control districts except for the number of births (see section 6.3 and Table 6.A.1). Columns 3 and 4 in Table 6.2 display means of relevant socio-economic characteristics comparing the families from treated and control neighborhoods. It becomes apparent that the two groups exhibit significant differences in some key SES characteristics: Specifically, individuals in the treatment group are significantly more likely to have only a basic school degree, no professional training, to be unemployed (before birth), to have a lower income and to be younger, and less likely to be employed full-time.

In order to account for these differences, we combine our IV and reduced form estimations with entropy balancing (Hainmueller, 2012), a matching strategy that balances pre-treatment controls more effectively than comparable propensity score methods. This matching step is conducted before running the IV or reduced form estimations. The main idea of entropy balancing is to assign a weight to observations in the control group, causing the control group's distributions of the selected covariates to match those of the treatment group in the first two moments, i.e., on mean and variances. As a result, the selected covariates in the treatment and control groups have the same means and variances. If several weighting schemes fulfill this balancing criterion, entropy balancing chooses the weighting scheme where all weights are non-negative and deviate the least from uniform weights (Hainmueller, 2012).

Two-Stage Least Squares. Next, we apply our instrument in a 2SLS approach to estimate the causal effect of program participation. In the first stage, we regress the program participation variable that we assume to be endogenous on our instrument and the exogenous control variables:

$$PK_i = \gamma_1 + \gamma_2 T_i + X_i' \gamma_4 + \varepsilon_i \tag{6.2}$$

Table 6.2: Descriptive statistics

	Unit range	BRISE mean	Treatment mean	Control mean	Difference b
Pro Kind participation					
Participation (t1)	1/0	0.08	0.17	0.01	-0.16***
Participation (t2)	1/0	0.09	0.19	0.02	-0.17***
Control variables					
Mother: no school degree	1/0	0.09	0.12	0.06	-0.06
Mother: General school degree	1/0	0.35	0.42	0.31	-0.11*
Mother: Highschool degree	1/0	0.56	0.46	0.63	0.17**
Mother: no training	1/0	0.26	0.34	0.21	-0.13*
Mother: Apprenticeship degree	1/0	0.40	0.36	0.43	0.07
Mother: Academic degree	1/0	0.34	0.30	0.37	0.07
Mother: Not in the labor force	1/0	0.45	0.53	0.40	-0.13*
Mother: Working part-time	1/0	0.29	0.27	0.30	0.03
Mother: Working full-time	1/0	0.26	0.20	0.30	0.10*
Father: Not in the labor force	1/0	0.18	0.16	0.19	0.03
Father: Working part-time	1/0	0.14	0.15	0.13	-0.02
Father: Working full-time	1/0	0.68	0.69	0.68	-0.01
First child	1/0	0.56	0.59	0.53	-0.05
Both parents born in Germany	1/0	0.50	0.44	0.54	0.10
One parent born outside Germany	1/0	0.20	0.23	0.17	-0.06
Both parents born outside Germany	1/0	0.31	0.33	0.29	-0.04
Single mother	1/0	0.10	0.08	0.11	0.03
Risk pregnancy	1/0	0.32	0.30	0.34	0.04
Household net income	Euros	2586.93	2393.13	2723.47	330.35*
Age Mother	years	31.08	30.20	31.71	1.51*
Satisfaction with family situation	0-2	1.86	1.83	1.87	0.04
Outcomes					
Smoking (t1)	1/0	0.12	0.14	0.11	-0.03
Smoking (t2)	1/0	0.13	0.16	0.11	-0.06
Alcohol (t1)	1/0	0.09	0.09	0.09	-0.00
Alcohol (t2)	1/0	0.18	0.17	0.18	0.01
Breastfeeding (t1)	1-4	3.48	3.39	3.55	0.15
Breastfeeding (t2)	1-3	2.66	2.58	2.71	0.13
EPDS (t1)	0-30	6.65	6.90	6.48	-0.42
EPDS (t2)	0-30	5.92	6.26	5.69	-0.57
Change in living cond. (t1)	8-32	26.43	26.15	26.61	0.46
Change in living cond. (t2)	8-32	27.05	27.04	27.05	0.02
Soothing strategies (t1)	9-54	48.42	48.11	48.62	0.51
Soothing strategies (t2)	9-54	47.00	45.71	47.88	2.17***
MONDEY milestones (t1)	0-26	18.10	18.37	17.91	-0.46
MONDEY milestones (t2)	0-53	32.77	32.90	32.68	-0.23
N		300	124	176	300

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns two to four show the mean values of major SES characteristics for the whole BRISE sample and separately for the treatment and control group. Column five indicates whether the difference between the two is statistically significant. Column one indicates the unit rage of the respective variable. 1/0 indicates a binary variable. *Satisfaction with the family situation* can take values 0 (not satisfied) to 2 (very satisfied). *Breastfeeding* can take values 1 (has never breastfed), 2 (breastfed only in the fist 4 weeks), 3 (breastfed longer than 4 weeks but not anymore) to 4 (I still breastfeed) in t1, and values 1 (never breastfed), 2 (only in the fist 4 weeks), and 3 (still breastfeed) in t2. The variable *change in living conditions* is composed of eight separate items which can take values ranging from 1 (do not agree) to 4 (completely agree), e.g. "raising my child brings me joy". Soothing strategies comprises nine separate items, e.g. I carry my child around when they cry. The *MONDEY milestones* count the number of "milestones" reached at three (t1) and seven (t2) months, e.g. whether the child is able to hold an object.
Source: BRISE (2017-2020), own calculations.

where T_i equals one if the family lives in a treatment neighborhood, PK_i, and X_i' are defined as above in Equation 6.1. The first stage regression is estimated using weighted least squares using the weights from entropy balancing. Since the dependent variable is binary, this corresponds to a linear probability model (LPM). In the second stage, the fitted values of the linear probability model from the first stage $\widehat{PK}_i$ are included as the main explanatory variable:

$$y_i = \beta_1 + \beta_2 \widehat{PK}_i + X_i'\beta_3 + \mu_i \tag{6.3}$$

In this regression, y_i are the different child and maternal outcome variables described in section 6.4. X_i' is again our vector of control variables that is the same as in the first stage regression. β_2 is our coefficient of interest and reflects the 2SLS estimator. Per definition, it estimates the local average treatment effect (LATE)[24] and thus depicts the effect of program participation on our outcomes.[25]

6.5.2 Cost analysis

Cost-efficiency analyses compare the "input" and "output" of measures. On the input side, this means the costs incurred by the measures in question (Spieß, 2013). In our case, this requires analyzing the complete picture of the costs required to run the home visiting program *Pro Kind*. To date, there is no standardized approach to capture the costs of programs (Karoly, 2012; Schmitz and Kröger, 2017) but the most rigorous approach is the ingredients-based method (Levin and McEwan, 2000). This approach begins by collecting detailed information on the types and quantities of resources used and then goes on to attach market or shadow prices to these resources. This means that besides the more obvious cost factors such as personnel costs, it is necessary to also account for "indirect" costs such as in-kind resources used and the opportunity costs (Spieß, 2013).

We conducted yearly cost surveys following the ingredient method, sending out questionnaires to the German Red Cross in Bremen, which is the executing agency of the *Pro Kind* intervention. The surveys take place in a Pen-and-paper Personal Interview (PAPI) version annually between spring and autumn, and the contact persons are asked to provide retrospective estimates of the time allocations of *Pro Kind* employees in the previous calendar year. The cost survey comprises detailed questions on different cost items, such as non-administrative and administrative personnel costs, volunteer

[24]It measures the effect on the compliers, i.e., those families whose program participation is induced by the access and information treatment.
[25]The robust standard errors μ_i are clustered at the household level.

activities, material resources used, investment goods, capital costs, cost of contractual services from third parties, costs for buildings and premises, training measures, overheads, as well as on the planned and actual utilization of the offered services. The information provided by the respondents is then fed into a micro database and analyzed (*BRISE* cost database). The database relies on a 100% response rate of the providers. The current market price for each employee is determined by the collective agreement for the public service (*TVöD*). Material resources that last for multiple years, e.g., office supplies like printers and pedagogical resources like books and toys, are annualized over five years, i.e., their average lifetime. Similarly, professional training measures are averaged over five years.

6.6 Empirical results

6.6.1 Effectiveness analysis

Table 6.3 reports the OLS results, i.e., coefficients obtained by simply regressing the different maternal and child outcomes measured when the child is three and seven months old on the *Pro Kind* participation rate. Across all outcomes, the point estimates are small in size and statistically not significant.[26]

Next, we turn to the results of the 2SLS estimations using the access and information treatment assigned on the neighborhood level as an instrument for program participation (Table 6.4). The first column shows first stage coefficients, namely the effect of living in a treated neighborhood with easier access and information provision on the *Pro Kind* participation rate at three months. Column two shows the reduced form estimates, i.e., the effect of living in treated neighborhoods on child and maternal outcomes at 3 months independent of actual *Pro Kind* program take-up (intention-to-treat effect). Lastly, column three presents the IV estimates, namely the causal effect of participation on child and maternal outcomes at three months. This IV-estimate constitute a LATE-effect, i.e. it estimates the causal effect for the subgroup of compliers, that is, the effect of *Pro Kind* participation for families who participate in the program because of the information provision and easier access to *Pro Kind* but would not have done so otherwise. Columns (4) - (6) show the respective results when the child is seven months old. All regressions are estimated using entropy balancing weights such that control group's distributions of the selected covariates match those of the treatment group in the first two moments.

[26]Due to missing values in the outcome variables, our sample size diminishes from the original 300 observations to 168-259 observations, depending on the outcome and specification.

Table 6.3: OLS Results

	3 months	7 months	3 months	7 months
	Maternal smoking		Maternal alcohol consumption	
Pro Kind	0.001	0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	247	205	244	198
	Breast feeding		Postnatal depression	
Pro Kind	-0.003	-0.002	0.001	-0.002
	(0.004)	(0.003)	(0.003)	(0.003)
Observations	249	168	259	259
	Change in living conditions		Soothing strategies	
Pro Kind	-0.001	0.005	0.003	-0.002
	(0.003)	(0.004)	(0.004)	(0.004)
Observations	229	172	232	198
	Child development: MONDEY score			
Pro Kind	0.005^{+}	0.001		
	(0.003)	(0.003)		
Observations	235	200		

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions include the full set of individual control variables. Robust standard errors in parentheses.
Source: BRISE (2017-2020), own calculations.

The first stage coefficients are highly significant and of similar magnitude across specifications: Living in a treatment neighborhood increases the *Pro Kind* participation rate by about 12 to 18 percentage points (depending on the outcome and age of the child). This suggests that the easier access and the information on this program provided to the parents in treated neighborhoods successfully nudged families to participate in *Pro Kind* .

The first panel shows the results for maternal smoking behavior. The reduced form and IV estimates are positive, but very imprecisely estimated: The size of the coefficients in column (3) and (6) amount to a 0.1 percentage point increase in maternal smoking at three months and a 0.2 percentage point increase at seven months if *Pro Kind* participation rates are increased by 10 percentage points.

Next, the second panel displays the results on maternal alcohol consumption. Again, we do not detect any significant effects. Coefficients are negative when the child is three months old, suggesting maternal consumption might slightly decrease in response to the treatment/*Pro Kind* participation. In contrast, coefficients turn positive and are of similar magnitude when children are seven months old. An opposing picture is visible for breastfeeding behavior: positive coefficients at three months and negative coefficients at seven months. As for the other outcomes, the effects are insignificant due to the small sample size and too imprecisely estimated to draw any conclusions.

Furthermore, there are also no significant effects on maternal postnatal depression and mothers' perception of the change in living conditions. However, also these results are too imprecisely estimated to draw conclusions. The last outcome concerning maternal behavior is the index characterizing the frequency of applying soothing strategies. Here, we depict negative and insignificant effects at three months and negative and statistically significant effects at seven months. Thus, the results suggest that living in the treatment districts decreases the application of soothing strategies by about 0.4 standard deviations and participating in *Pro Kind* by 0.024 standard deviations.

Finally, the last panel shows the results of the child development indicator MONDEY. Coefficients are positive at three months and negative at seven months but statistically not significant. We can establish with 95% certainty that the IV effects are not larger than 0.02 and not smaller than -0.023 standard deviations for both measuring times. Thus, despite the small sample size, we can conclude that *Pro Kind* participation – at least in the very short run – does not substantially impact child development.[27]

The results in our preferred specification are based on the full sample (including first and not first time mothers) and the continuous definition of *Pro Kind* participation and employing entropy balancing techniques. Additionally, we employ a binary definition, i.e., turning one if the participation rate is at least 50%. The results presented in Table 6.A.4 are very similar to our main results, i.e., displaying insignificant effects across all outcomes except for soothing strategies at seven months. Only first time mothers are eligible for *Pro Kind participation*. Since in our sample only a bit more than half of all mothers are first time mothers, a significant share of the treatment group (living in treatment districts) are not eligible for *Pro Kind*, and thus by definition non-compliers. In our main specification we use the full sample to increase the sample size. Table 6.A.5 reports the results based on a sample restricted to first time mothers. As expected, the first stage coefficients increase: Living in a treatment district increases program participation by 21 to 32 percentage points. However, the coefficients obtained by the reduced form and IV estimations remain statistically insignificant for all outcomes.

[27]Regressions on the separate MONDEY domains and items yielding the "Mother's perception of change in living conditions" and "Soothing strategies" indices also do not yield statistically significant estimates. Results are available upon request.

Table 6.4: IV Results

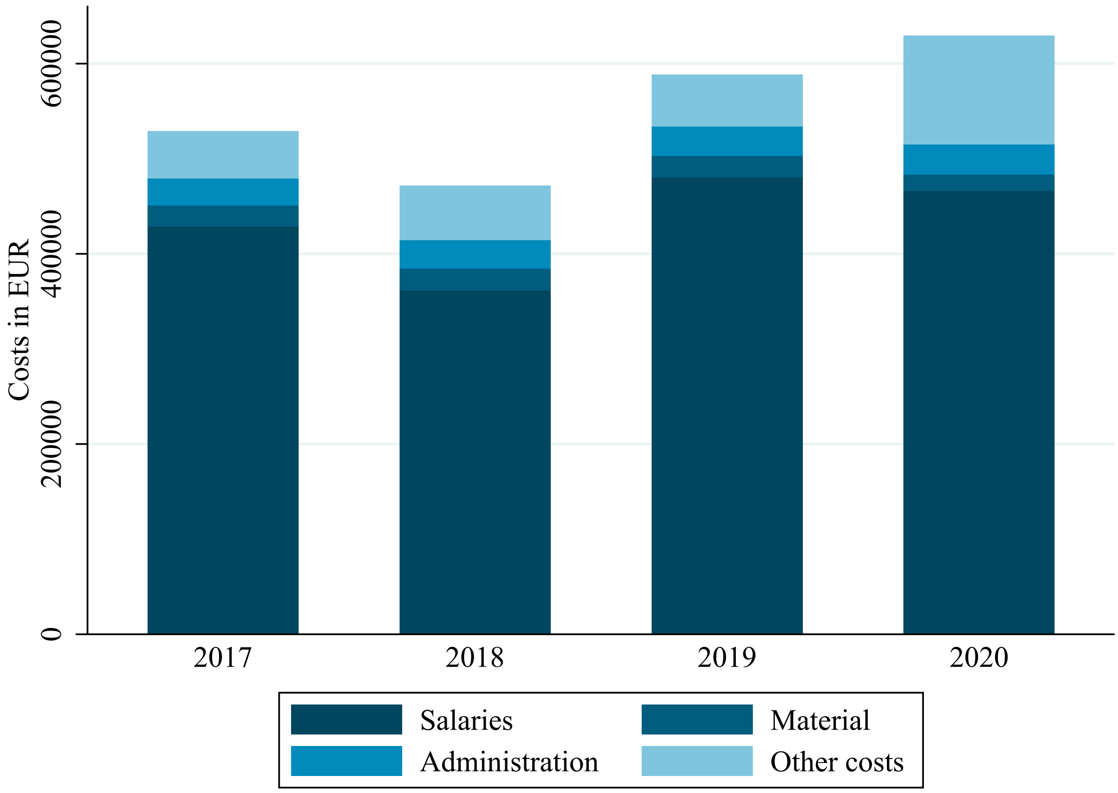
	First stage	Reduced form	IV	First stage	Reduced form	IV
	Maternal smoking behavior					
	3 months			7 months		
Info Treatment	11.574***	0.013		13.866***	0.035	
	(2.849)	(0.045)		(3.392)	(0.049)	
Pro Kind			0.001			0.002
			(0.004)			(0.003)
Observations	247	247	247	205	205	205
$F-statistic$			18.732			19.462
	Maternal alcohol consumption					
	3 months			7 months		
Info Treatment	11.930***	-0.010		14.997***	0.011	
	(2.901)	(0.041)		(3.685)	(0.060)	
Pro Kind			-0.001			0.001
			(0.003)			(0.004)
Observations	244	244	244	198	198	198
$F-statistic$			19.214			18.903
	Breast feeding					
	3 months			7 months		
Info Treatment	11.564***	0.008		15.240***	-0.063	
	(2.825)	(0.141)		(4.199)	(0.167)	
Pro Kind			0.001			-0.004
			(0.011)			(0.010)
Observations	249	249	249	168	168	168
$F-statistic$			18.978			14.727
	Maternal postnatal depressions					
	3 months			7 months		
Info Treatment	12.175***	0.073		14.110***	0.134	
	(2.773)	(0.131)		(2.981)	(0.129)	
Pro Kind			0.006			0.009
			(0.010)			(0.009)
Observations	259	259	259	259	259	259
$F-statistic$			21.766			24.978
	Mother's perception of change in living conditions					
	3 months			7 months		
Info Treatment	11.888***	-0.081		17.643***	-0.081	
	(3.117)	(0.139)		(4.033)	(0.158)	
Pro Kind			-0.006			-0.005
			(0.011)			(0.009)
Observations	229	229	229	172	172	172
$F-statistic$			16.518			19.943
	Soothing strategies					
	3 months			7 months		
Info Treatment	11.754***	-0.035		15.920***	-0.398**	
	(3.030)	(0.154)		(3.653)	(0.152)	
Pro Kind			-0.003			-0.024*
			(0.012)			(0.010)
Observations	232	232	232	198	198	198
$F-statistic$			16.753			21.088
	Child development: MONDEY score					
	3 months			7 months		
Info Treatment	11.614***	-0.015		16.315***	0.014	
	(3.089)	(0.138)		(3.846)	(0.153)	
Pro Kind			-0.001			0.001
			(0.011)			(0.009)
Observations	235	235	235	200	200	200
$F-statistic$			15.704			20.457

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. All regressions include the full set of individual control variables. Regressions are weighed with entropy balancing weights.
Source: BRISE (2017-2020), own calculations.

6.6.2 Program costs

Figure 6.1 displays the mean costs of *Pro Kind* based on our detailed yearly cost surveys (2017 - 2020). The costs refer to the total costs of *Pro Kind* in the city of Bremen, of which an increasing share participates in *BRISE*.[28] Personnel costs constitute by far the largest cost share, making up around 80 percent of the total costs. This is in line with similar programs, which are generally labor intensive (e.g., Workman, 2018). In 2020, the *Pro Kind* staff had to move offices, which generated above average "other costs". Table 6.A.3 in the Appendix provides a more detailed breakdown of the costs.

Figure 6.1: Cost development *Pro Kind*

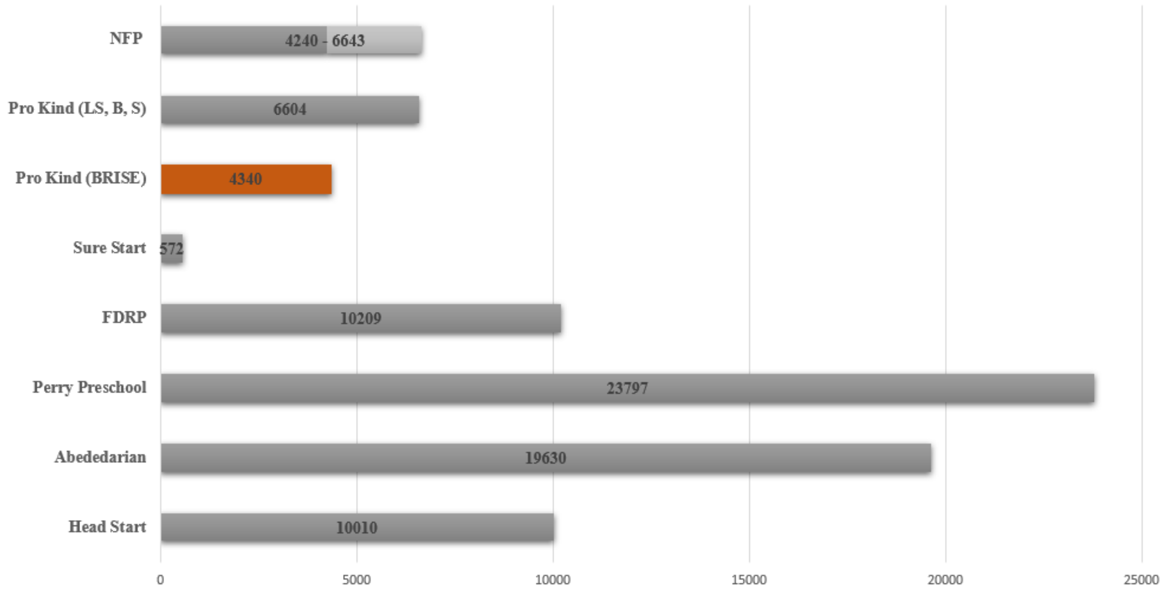


Notes: Costs in Euro per year.
Source: BRISE 2017-2020, own calculations.

For the years 2017 through 2020, the average costs per participant per year in *Pro Kind* ranged between 3,468 and 3,861 Euros, and costs per home visit ranged between 265 and 325 Euros (Table 6.5). Figure 6.2 relates the total cost per participant per year to other early childhood programs that also entail parenting support elements. The comparison reveals that *Pro Kind* in Bremen is less expensive than the average

[28]While in 2017, only five BRISE children participated in *Pro Kind*, this number rose to 63 in 2020.

Figure 6.2: Comparison of per child costs of selected programs in 2021 USD



Notes: Costs are per child per year and are inflated to 2021 USD. The cost estimates may refer to different data collection methods, due to lack of a harmonized approach (Karoly, 2012).
Sources: Head Start: 7000$ in 2004 USD (U.S. Department of Health & Human Services, 2004); Abecedarian project: 13000$ in 2002 USD (Barnett and Masse, 2007); Perry Preschool: 17759 $ in 2006 USD (Heckman et al., 2010); FDRP: 7345$ in 2005 USD (Besharov et al., 2011); Sure Start: 416GBP in 2021 GBP (Cattan et al., 2021); *Pro Kind* (Brise): €3669 in 2021 EUR (BRISE, own calculation); *Pro Kind* (LS= Lower Saxony, B= Bremen, S= Saxony): €4353 in 2012 EUR (Maier-Pfeiffer et al., 2013); NFP: 3420-5358$ in 2010 USD (Miller and Hendrie, 2015).

cost estimated by Maier-Pfeiffer et al. (2013) for *Pro Kind* in three German federal states. In fact, *Pro Kind* within BRISE has lower average costs per child than any comparable program except Sure Start, which is a less intensive non-targeted program. The prominent Perry Preschool program is roughly 5.5 times as expensive as *Pro Kind*.

Table 6.5: Pro Kind cost summary: 2017-2020

	2017	2018	2019	2020
Costs per participant	3861.2	3467.6	3644.4	3702.7
Costs per home visit	324.53	304.1	309.9	264.815

Notes: Costs are reported in 2021 EUR.
Source: BRISE (2017-2020), own calculations.

6.7 Conclusion

Investments in early childhood have the potential to effectively mitigate socio-economic inequalities (e.g., Heckman et al., 2010). While center-based daycare programs play an important role (e.g., Barnett, 1985; Havnes and Mogstad, 2011b; Heckman et al., 2013; Karoly et al., 2006; Kautz et al., 2014), home-visiting programs targeting parenting skills and knowledge are an equally important – and in the European context under-researched – component of early childhood education and care. These programs advise parents on aspects of everyday life with an infant, such as mother-and-child interaction, nutrition, and ways to seek support when needed. An often neglected aspect of impact evaluations is the cost of these programs. Comprehensively assessing both the benefits and costs of investments builds the basis for later cost-effectiveness and cost-benefit analyses.

In this study, we present novel evidence on the early effects and costs of the home-visiting program *Pro Kind* under the Bremen Initiative to Foster Early Childhood Development (BRISE). We find no economically meaningful or statistically significant effects of *Pro Kind* on an extensive range of outcomes covering different aspects related to maternal and child well-being at three and seven months after childbirth. To establish causality, we exploit a random information and access treatment to this program at the neighborhood level that induced families to participate in *Pro Kind*.

Our cost analysis reveals that the average costs per participant per year in *Pro Kind* range between 3,468 and 3,861 Euros over the study period (2017–2020). In international comparison, *Pro Kind* belongs to the less costly programs, with its cost corresponding to about 18 percent of the per-child-per-year expenditure of the Perry Preschool program.

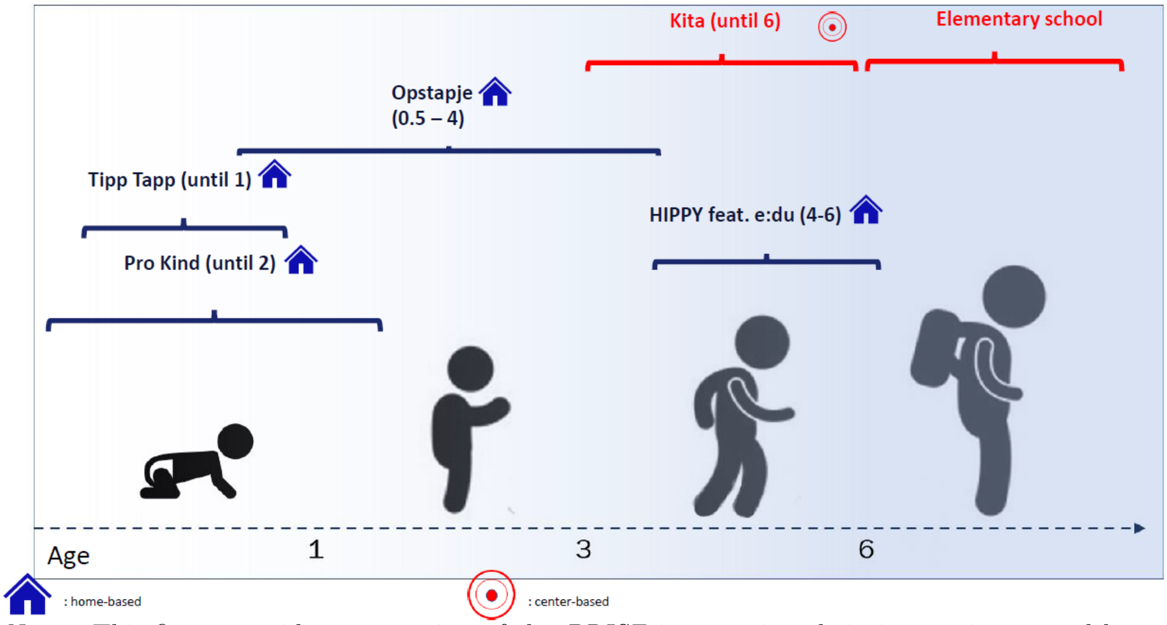
Our findings regarding *Pro Kind*'s effectiveness in improving mother and child outcomes should be interpreted in the light of the data challenges we faced: With just 300 total observations, we have a relatively small sample, which is further diminished by missing values in important control variables. This forces us to impute variables in some cases (e.g., household income) and to exclude other potentially important variables, such as the father's education. This leads to imprecisely estimated coefficients which do not allow us to derive conclusions regarding the early effects of *Pro Kind*. Lastly, randomization of the treatment on the neighborhood level was non-perfect as families significantly differ in some socio-economic characteristics (Table 6.2). We tackle this problem by employing entropy balancing. However, perfect randomization is always cleaner than imposed randomization, especially given the incomplete set of control variables.

Furthermore, it is not surprising that *Pro Kind* does not significantly impact child and mother outcomes in a child's first months. It usually takes several years for the effects of such programs to materialize (e.g., Jungmann et al., 2015; Sandner, 2013, 2019; Sandner et al., 2018). Hence, future studies evaluating the effectiveness of *Pro Kind* at later stages are better positioned to conduct a sound analysis. Our analysis paves the way for later cost-effectiveness and cost-benefit analyses within *BRISE*. At later measurement times and with a larger sample size, it should be possible to not only achieve more meaningful effectiveness estimations for *Pro Kind* but also to measure the effects of *Opstapje* - another program that starts shortly after birth and is part of the chain of interventions set up within BRISE. Along with the yearly collected cost data for all programs within *BRISE*, this allows to conduct a cost-effectiveness study, which is only possible when comparing at least two separate programs (Karoly, 2012; Spieß, 2013).

6.A Appendix

6.A.1 Additional information on *BRISE*

Figure 6.A.1: *BRISE* intervention chain



Notes: This figure provides an overview of the *BRISE* intervention chain integrating several home-based and center-based programs from birth to school.

Source: *BRISE* consortium 2022.

Figure 6.A.2: Pro Kind Flyer

Notes: This figure depicts the information flyer for the program *Pro Kind* provided to *BRISE* families by the family counselors.

Source: *BRISE* consortium.

6.A.2 Additional results

Table 6.A.1: Randomization by city districts

	Control group mean	Treatment group mean	Difference b
School degree	83.82	84.16	-0.342
Unemployment transfers	34.68	40.61	-5.928
Relocation from district	35.41	30.43	4.976
Share of inhabitants with migration background	19.78	23.17	-3.394
Birth rate	100.98	110.64	-9.658
Number of births	72.12	95.40	-23.282*
Complete social Index	537.550	572.69	-35.137
N	17	10	27

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

Source: Administrative data from the city of Bremen, own calculations.

Table 6.A.2: Descriptive statistics: Brise vs. SOEP

	BRISE mean	SOEP mean
Mother: no school degree	0.09	0.01
Mother: Basic school degree	0.35	0.48
Mother: Highschool degree	0.56	0.51
Mother: no training	0.26	0.18
Mother: Apprenticeship degree	0.40	0.50
Mother: Academic degree	0.34	0.32
Mother: Not in the labor force	0.45	0.22
Mother: Working part-time	0.29	0.44
Mother: Working full-time	0.26	0.34
Father: Not in the labor force	0.18	0.19
Father: Working part-time	0.14	0.14
Father: Working full-time	0.68	0.67
No migration background	0.50	0.70
One parent with migration background	0.20	0.19
Both parents with migration background	0.31	0.11
Household net income	2586.93	3818.75
Age Mother	31.08	31.91
Single mother	0.10	0.16
Satisfaction with family situation	1.86	1.83
First child	0.56	0.34
N	300	494

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The first two columns show the mean values of major SES characteristics for the treatment and control group. Column three indicates whether the difference between the two is statistically significant. Column four shows the mean value for the same variables in the SOEP, restricting the sample to urban households, families with children below one and dropping the SOEP migration (M1-M5), low-income (L2) samples and all observations surveyed before 2010.
Source: BRISE (2017-2020), SOEP v.36 (2010-2019), own calculations.

Table 6.A.3: Pro Kind cost summary: 2017-2020

Variable	Obs	Mean	Std. Dev.	Min	Max
Resources used					
Labor costs (excl. admin.)					
Number of employees (excl. administration)	4	16.75	1.5	15	18
Weekly hours employees (without administration)	4	306.356	45.582	246.766	357.678
Weekly hours per employee	4	18.24	1.589	16.451	19.871
Total personnel costs (excl. admin.)	4	426493.3	45843.69	361623	466005
Wage per employee	4	25507.61	2460.844	23823.69	29125.31
Personnel costs (admin.)					
Number of employees (administration)	0	.	.	.	.
Weekly hours employees (administration)	0	.	.	.	.
Total personnel costs (admin.)	4	30266.51	1444.765	28470.46	31898.64
Material costs					
Total value of expenses material	4	21268.81	2674.629	17274.39	22808.41
Capital goods					
Expenditure on capital goods	4	.75	.5	0	1
Purchase price of capital goods	2	4792.655	469.766	4460.48	5124.83
Rental/lease expenses capital goods	1	615.96	.	615.96	615.96
Amount of capital costs	3	601.317	313.288	247.22	842.48
Costs for rooms					
Rental costs for rooms	4	26425.96	7951.809	14995.05	32745.59
Other ressources					
Costs for other resources	4	16556.56	11006.9	9819.29	32992.46
Costs for further training	4	10009	3589.392	4741.65	12787.69
Overheads					
Total overheads (excl. admin.)	4	30266.51	1444.765	28470.46	31898.64
Use of the offers					
Planned utilization					
Places with committed funding (planned)	4	138.75	2.5	135	140
Home visits per family (planned)	4	22	0	22	22
Actual utilization					
Places used (realized)	4	149	16.021	136	170
Number of regularly participating families	2	54	16.971	42	66
Number of families who joined later	2	107.5	4.95	104	111
Total home visits (realized)	4	1839.25	373.126	1551	2377
Home visits per family (realized)	2	22	0	22	22
Average number of home visits (reg.)	2	18.765	1.082	18	19.53
Average number of home visits (irreg.)	2	10.23	.325	10	10.46
Number of home visits by phone, videocall, walk	1	1221	.	1221	1221
General information					
Total costs in Bremen	4	546908.1	65633.71	471592.5	629464.4
Total costs within BRISE	4	141128.2	99852.28	19306	233272.1
Personnel as a share of total costs	4	.795	.053	.74	.861
Proportion of admin. in total costs	4	.056	.005	.051	.063
Share of materials in total costs	4	.04	.009	.027	.048
Share of other costs in total costs	4	.124	.041	.093	.182
Costs per participant	4	3668.978	162.532	3467.592	3861.201
Costs per home visit	4	300.837	25.51	264.815	324.53

Notes: The table displays mean, standard deviation, minimum and maximum of costs reported in EUR.
Source: BRISE (2017-2020), own calculations.

6.A.3 Robustness

Table 6.A.4: IV Results: Treatment-dummy

	First stage	Reduced form	IV	First stage	Reduced form	IV
	Maternal smoking behavior					
	3 months			7 months		
Info Treatment	0.121***	0.013		0.150***	0.035	
	(0.034)	(0.045)		(0.043)	(0.049)	
Pro Kind			0.098			0.219
			(0.336)			(0.301)
Observations	247	247	247	205	205	205
F − statistic			14.264			13.565
	Maternal alcohol consumption					
	3 months			7 months		
Info Treatment	0.125***	-0.010		0.175***	0.011	
	(0.035)	(0.041)		(0.047)	(0.060)	
Pro Kind			-0.075			0.060
			(0.301)			(0.311)
Observations	244	244	244	198	198	198
F − statistic			14.685			15.265
	Breast feeding					
	3 months			7 months		
Info Treatment	0.121***	0.008		0.165**	-0.063	
	(0.034)	(0.141)		(0.050)	(0.167)	
Pro Kind			0.062			-0.360
			(1.056)			(0.913)
Observations	249	249	249	168	168	168
F − statistic			14.402			12.244
	Maternal postnatal depressions					
	3 months			7 months		
Info Treatment	0.126***	0.073		0.150***	0.126	
	(0.033)	(0.131)		(0.037)	(0.121)	
Pro Kind			0.547			0.826
			(0.961)			(0.803)
Observations	259	259	259	259	259	259
F − statistic			16.209			18.159
	Mother's perception of change in living conditions					
	3 months			7 months		
Info Treatment	0.126***	-0.081		0.194***	-0.081	
	(0.037)	(0.139)		(0.051)	(0.158)	
Pro Kind			-0.594			-0.409
			(0.988)			(0.779)
Observations	229	229	229	172	172	172
F − statistic			13.278			14.558
	Soothing strategies					
	3 months			7 months		
Info Treatment	0.124***	-0.035		0.174***	-0.398**	
	(0.036)	(0.154)		(0.045)	(0.152)	
Pro Kind			-0.263			-2.212*
			(1.133)			(0.960)
Observations	232	232	232	198	198	198
F − statistic			13.296			15.816
	Child development: MONDEY score					
	3 months			7 months		
Info Treatment	0.122***	-0.015		0.180***	0.014	
	(0.036)	(0.138)		(0.048)	(0.153)	
Pro Kind			-0.112			0.075
			(1.039)			(0.786)
Observations	235	235	235	200	200	200
F − statistic			12.536			15.651

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. All regressions include the full set of individual control variables. Regressions are weighed with entropy balancing weights.
Source: BRISE (2017-2020), own calculations.

Table 6.A.5: IV Results: First-time mothers

	First stage	Reduced form	IV	First stage	Reduced form	IV
	Maternal smoking behavior					
	3 months			7 months		
Info Treatment	21.106***	0.029		28.244***	0.027	
	(4.848)	(0.053)		(5.695)	(0.060)	
Pro Kind			0.001			0.001
			(0.002)			(0.002)
Observations	137	137	137	110	110	110
F − statistic			19.561			25.783
	Maternal alcohol consumption					
	3 months			7 months		
Info Treatment	21.624***	-0.006		30.555***	0.070	
	(4.935)	(0.040)		(6.249)	(0.082)	
Pro Kind			-0.000			0.002
			(0.002)			(0.002)
Observations	136	136	136	106	106	106
F − statistic			19.795			25.082
	Breast feeding					
	3 months			7 months		
Info Treatment	21.324***	-0.051		30.966***	-0.053	
	(4.853)	(0.203)		(7.474)	(0.209)	
Pro Kind			-0.002			-0.002
			(0.009)			(0.006)
Observations	137	137	137	89	89	89
F − statistic			19.827			18.257
	Maternal postnatal depressions					
	3 months			7 months		
Info Treatment	21.781***	-0.106		27.897***	0.113	
	(4.719)	(0.166)		(5.016)	(0.169)	
Pro Kind			-0.005			0.004
			(0.007)			(0.006)
Observations	142	142	142	142	142	142
F − statistic			21.596			31.278
	Mother's perception of change in living conditions					
	3 months			7 months		
Info Treatment	22.062***	-0.340+		31.363***	-0.058	
	(5.450)	(0.179)		(6.413)	(0.211)	
Pro Kind			-0.014+			-0.002
			(0.008)			(0.006)
Observations	127	127	127	99	99	99
F − statistic			17.771			25.470
	Soothing strategies					
	3 months			7 months		
Info Treatment	20.663***	0.110		29.587***	-0.439*	
	(5.160)	(0.232)		(6.063)	(0.212)	
Pro Kind			0.005			-0.014*
			(0.010)			(0.007)
Observations	128	128	128	112	112	112
F − statistic			16.434			24.085
	Child development: MONDEY score					
	3 months			7 months		
Info Treatment	22.189***	0.064		31.802***	-0.093	
	(5.287)	(0.211)		(6.330)	(0.219)	
Pro Kind			0.003			-0.003
			(0.008)			(0.006)
Observations	132	132	132	110	110	110
F − statistic			18.129			25.041

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. All regressions include the full set of individual control variables. Regressions are weighed with entropy balancing weights.
Source: BRISE (2017-2020), own calculations.

CHAPTER 7

Conclusion

Health and well-being are essential for individuals across all generations and across society as a whole. Social policies targeting a particular domain, for example, human capital accumulation, often also impact health. In this dissertation, I contribute to understanding the (unintended) consequences of social and family policies on the health and well-being of different generations. Specifically, this dissertation comprises five chapters examining how selected social and family policies in the German context impact health, well-being, and associated costs. This final chapter draws a conclusion by briefly discussing the policy implications and limitations of the studies[1] and by highlighting avenues for future research.

Chapters 2 and 5 investigate the impact of an increase in the early retirement age for women by three years on individuals' health (**Chapter 2**) and associated healthcare costs (**Chapter 5**). The results indicate that prolonging the working life leads to worse health outcomes. Specifically, the prevalence of mental disorders, musculoskeletal diseases, and obesity increases, while we do not observe improvements for any of the other diagnoses under study. In line with an increased prevalence of some diagnoses, we also find an increase in healthcare consumption and costs. However, when contrasting the increase in costs to the positive fiscal effects of the pension reform, it turns out that the costs amount to only about 2% of the generated revenues.

In light of an aging society, future reforms to the retirement age seem inevitable to sustain the financial stability of the public pension system. Taking an isolated health perspective, from **Chapter 2** one could imply that future increases in the retirement age should be avoided. However, the comparison of the revenues of the

[1]A more thorough discussion of policy implications and limitations can be found in the individual chapters. Generally, results of empirical analyses rely on the validity of the underlying assumptions and are, to some degree, context-specific.

reform to the increase in healthcare costs in **Chapter 5** demonstrates that the reform had positive fiscal effects. Although individual costs of ill health and its indirect costs, such as productivity losses, are not included in this estimate, there is no doubt that, in general, the revenues exceed the costs of the reform. Our results, however, call for accompanying actions to compensate for the unintended consequences on health. Examples of such actions are preventive measures and health investments across the human life cycle that could build resiliency in the workplace. Furthermore, investing in education throughout the employment history could facilitate occupational changes for older workers. Likewise, working conditions must be adapted to age so that work strain, which increases with age, can be accommodated. Policymakers must also ensure that the invalidity pension covers health risks for people who can no longer work. Thus, any further increase in the retirement age must be flanked by additional reforms to the invalidity pension.

A limitation of the findings is that **Chapters 2 and 5** estimate intention-to-treat effects since the data neither include information on the working history of women nor their eligibility for early retirement (old-age pension for women). From Geyer and Welteke (2021) and Geyer et al. (2020), we know that 60% of all women born in 1951 were eligible for the old-age pension for women and the abolishment of this retirement option mostly led to women staying in their employment status (e.g., in employment, unemployment or inactivity). However, we are not able to disentangle the mechanisms that drive the health effects; i.e., we do not know whether the health effects are homogeneous across women in different employment statuses or whether effects are, for example, more pronounced for the employed. For future research, it would be very interesting to disentangle the underlying mechanisms that drive the health effects. Another limitation of the studies is that we are not able to assess whether the health effects differ by socio-economic status. The KBV data do not include socio-economic characteristics on the individual level, except for the birth year, birth month, and gender. Since there is evidence that the health effects of retirement are heterogeneous across socio-economic groups (see, e.g., Etgeton and Hammerschmid, 2019, and references therein), it would be interesting to assess whether this also holds in this context.

Many findings of these chapters will apply to other countries with similar institutional settings (e.g., universal healthcare systems); still, our analyses come with limits to generalizability for other settings. For example, our results are derived from a pension reform solely affecting women. *Ex-ante*, it is unclear whether the effects would be similar for men as, on average, the employment histories for men and women of this generation differ substantially (OECD, 2022a). Furthermore, the reform induced a change in the retirement age by three years, which is relatively large compared to other settings. In contrast, for example, in recent times the statutory pension age in Germany was raised stepwise in (two)-months increments per birth cohort from 65 to 67 years (DRV, 2022). Thus, for future research, it could be interesting to assess whether the health effects are also present when the pension age is increased in smaller increments (e.g., monthly), ultimately with the aim of finding an optimal strategy to increase the retirement age while minimizing the negative health implications.

Promoting the health of the elderly within a society is important for many reasons. Besides individual benefits and savings to the healthcare system, the elderly (and particularly grandparents) are connected to other generations through family ties including, for example, care provision to grandchildren. The intergenerational ties through care provision highlight the relevance of assessing the effects of grandparental care on parental and child outcomes. **Chapter 3** contributes to a double-generation perspective and analyzes the effects of grandparental childcare on parental well-being and child outcomes. We find that grandparental care is particularly valuable for parents as it increases their satisfaction with the childcare situation and, for mothers, their satisfaction with their leisure time. In contrast, our results show that grandparental care does not affect most of the considered child outcomes. If any, we find adverse effects on children's health, driven mainly by children cared for by less healthy grandparents. Following parental care and daycare or school, grandparents act as the third biggest caregiver for children below ten. Even in light of recent expansions of (full-time) spots in daycare and increasing availability of all-day schools, grandparents remain important in the "care puzzle" of many families (Barschkett et al., 2022a). Since child and family outcomes are affected by parental care, daycare, and childcare provided by grandparents, policymakers should shift their focus with respect to both formal and informal childcare modes. For example, policymakers could consider introducing national insurance credits contributing to the pension income for grandparents providing care to their grandchildren, as implemented in the UK. Alternatively, policymakers could discuss grandparental leave and benefits options similar to the parental leave system, as in Portugal (Milovanska-Farrington, 2021).

However, our results also provide evidence that too many caregivers in one day can disrupt child development. Policymakers could address this issue in different ways. First, an increase in the share of full-time employed daycare teachers is desirable to provide stable and high-quality care in daycare centers. More full-time employed daycare teachers would reduce the number of caregivers a child experiences during a day in daycare. To attract more full-time employees to daycare centers, the working conditions must be improved, e.g., with higher wages. Second, in light of the increasing attendance of young children in daycare and constantly high rates of grandparental care, grandparents could be integrated into the daycare environment by, for example, offering "grandparents-days."[2] Thereby the different caregivers of a child would get to know each other, providing the basis for an effective contribution to the upbringing of a child. Third, policymakers could set the general framework for more family-friendly work environments that allow parents to reconcile childcare and paid work without relying on regular grandparental care. Examples of this are longer daycare opening hours and more flexible working schedules. Lastly, one way to improve outcomes of children in grandparental care could be an improvement in the quality of the provided care. With the data at hand, we are not able to assess the quality of care, measured by, for example, activities that grandparents conduct with the child. A promising avenue for future research is to find ways to measure the quality of care (for example based on time-use survey data), which could then be translated into policy recommendations that aim to improve the quality of care.

The findings of the chapter are highly relevant to the German context, but grandparental care varies strongly internationally. The underlying reasons for this are manifold; for example, differences in daycare systems, maternal labor market participation, and social norms are contributing factors. Thus, the external validity of our findings is limited to countries with similar institutional contexts. In terms of internal validity, a limitation of this study is our instrumental variable – distance to the grandparents – which we argue to be exogenous. We conduct extensive and careful robustness checks that provide evidence for the plausibility of the exogeneity assumption. Despite the reassuring results of the robustness checks, there are reasons to believe that our instrument is, in parts, endogenous. For example, parental well-being could contribute to the decision to live close to the grandparents. However, the possibilities for other sources of exogenous variation to establish causality between grandparental care and the parental and child outcomes are limited. Thus, our approach yields reasonably causal estimates under the constraint of current data availability and econometric methods.

Germany is characterized by a multi-actor childcare system, in which grandparents are the third biggest caregiver. Daycare (or school for older children) is the second biggest care actor, emphasizing the relevance of assessing the implications of this childcare environment. Consequently, **Chapter 4** estimates the effects of early daycare attendance on children's health. Specifically, I evaluate a daycare expansion for children below the age of three, leading to an earlier daycare entry age. The findings highlight a substitution of illness spells: In more detail, children who enter daycare below the age of three suffer from more communicable diseases at daycare entry age and

[2] An example could be an open-door day where children bring their grandparents to daycare.

are less often afflicted in elementary school. These effects are particularly strong for children from deprived areas. A similar pattern is visible for healthcare consumption. For the remaining outcomes, i.e., mental disorders, obesity, vision problems, injuries, and healthcare costs, I do not find evidence for an association with daycare attendance. While the substitution of illness spells is interesting, *per se*, it raises the question of whether it is beneficial. In the study, I discuss several potential implications, including cost effects, spill-over effects on parents and siblings, the duration of illness spells at different ages, and sickness absence at school or daycare. My considerations lead to the conclusion that from a health perspective, there is no evidence against an early daycare entry age. Hence, an obvious policy implication is expanding daycare slots for children below three such that the supply satisfies the demand. A plethora of studies provides evidence for positive effects of daycare on children's (non-)cognitive development, particularly for children from disadvantaged background (e.g., Cornelissen et al., 2018), and other relevant factors such as maternal satisfaction (Schmitz, 2020), maternal labor market participation (e.g., Müller and Wrohlich, 2020), and fertility (e.g., Bauernschuster et al., 2016). Furthermore, in the federal legislation on daycare, the government has identified several fields of action to improve daycare quality – one of these fields is child health.[3] Although child health is identified as an important field of action, to date, none of the 16 states have used federal funds to invest in this field of action. My findings call for action in this regard to make daycare centers and schools a safer place for children. For instance, policymakers could invest in smaller daycare groups and larger outdoor areas, which have both long been known to reduce the risk of infections (e.g., Grosch and Niebsch, 1987; Pönkä et al., 1991).

A limitation of this study is that the considerations regarding the implications of the results are mostly based on related findings from the international literature (except for an additional correlational analysis on spill-over effects on parents based on the SOEP), which highlights the scope for future research. Having established the substitution of illness spells for children entering daycare at age one or two, it would be interesting to expand the analysis to parents and siblings. To the best of my knowledge, there is no study looking at the causal effects of daycare attendance of children on parental health. Apart from Daysal et al. (2022), there is also limited evidence on spill-over effects on siblings. To fully understand the impact of daycare on family health, spill-over effects on parents[4] and siblings are important connecting factors to investigate. To do so, health data linking children, parents, and siblings with sufficient sample sizes are required, which are unavailable to date in Germany. Similarly, to better

[3]Gesetz zur Weiterentwicklung der Qualität und zur Teilhabe in der Kindertagesbetreuung vom 19. Dezember 2018. Bundesgesetzblatt. 2018;49:2696-9.

[4]Also potential spill-over effects on care-giving grandparents are of interest.

understand the health impact, it would be interesting to study the causal impact of daycare attendance on the duration of illness spells for the different age groups. With the data at hand, this is not possible: information on the duration of illness spells is missing. Another avenue for future research is investigating the implications of shifting illness spells on sickness absence at daycare and school. Sickness absence at school is associated with worse educational and labor market outcomes (e.g., Cattan et al., 2017), implying that reduced sickness absence at school could further benefit children. Lastly, the KBV data do not allow for conducting heterogeneity analysis on the individual level, which could provide further interesting insights into which group is particularly affected. My analysis by the level of average household income and share of migrant in the area where children live provides suggestive evidence for more pronounced effects for children from disadvantaged backgrounds. If this finding also holds true on the individual level, policymakers can target their actions more precisely on the most vulnerable groups.

Like **Chapter 3**, the findings are very relevant to inform the German policy debate on childcare modes, but the findings are only applicable in comparable institutional contexts, such as countries with a similar daycare system. While evidence from Sweden generates similar findings (van den Berg and Siflinger, 2022), evidence from Canada on a low-cost daycare expansion (e.g., Baker et al., 2008) and findings for targeted non-universal programs (e.g., Conti et al., 2016) exhibit different results. Hence, it is important to consider, among others, the daycare quality, the target group, and the studied outcomes in the specific country, when drawing cross-country comparisons.

Lastly, I turn the focus on the most important care-giver – the parents. In **Chapter 6**, we examine how *Pro Kind*, a parenting program entailing home visits by trained nurses within *BRISE*, affects short-term maternal and child developmental, behavioral and health outcomes. Additionally, we build a detailed micro cost database on program costs and analyze the different cost components. Our results do not provide evidence for effects at this early stage of a child's life. The cost analysis reveals that *Pro Kind* is one of the cheaper programs among comparable programs in other countries. It is not surprising that *Pro Kind* does not significantly impact child and maternal outcomes in the very early stages of a child's life, as evidence on *Pro Kind* in another context shows that effects usually take several years to materialize (e.g., Jungmann et al., 2015; Sandner, 2013, 2019; Sandner et al., 2018). Furthermore, our study faces severe data challenges. Our sample, with only 300 total observations, is small and further diminished by missing values in core variables. Additionally, the randomization within the trial was non-perfect, leading to differences in socio-economic characteristics between treatment and control families. Therefore, the current data does not allow us to estimate precise effects.

Despite our data challenges, our study builds a promising basis for future cost-effectiveness studies. At later measurement times and with larger sample sizes, the study setting allows us to derive more meaningful estimates of the program's effectiveness. Furthermore, later measurement times will also allow us to derive estimates for other ECEC programs within *BRISE*. Together with the detailed cost data on each program, a comparison of the programs allows for conducting a cost-effectiveness analysis. Cost-effectiveness studies are highly relevant for informing the policy debate, as they can guide policymakers' decisions between certain programs. For example, if *Pro Kind* turns out to positively impact child and maternal outcomes, the relatively low costs of *Pro Kind* compared to other programs makes the program an attractive option to support children from disadvantaged backgrounds, thereby leveling the playing field at an early stage in children's lives.

To conclude, with this dissertation, I contribute to the policy debates concerning various social and family policies affecting individual health and well-being. I derive policy recommendations that can help improving the health, well-being, and, thus, the lives of people of all generations while introducing interesting avenues for future research.

Bibliography

Aalto, A.-M., E. Mörk, A. Sjögren, and H. Svaleryd (2019): "Does childcare improve the health of children with unemployed parents? Evidence from Swedish childcare access reform," *IFAU Working Paper*, 2019/1.

Almond, D., J. Currie, and V. Duque (2018): "Childhood circumstances and adult outcomes: Act II," *Journal of Economic Literature*, 56, 1360–1446.

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

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of causal effects using instrumental variables," *Journal of the American statistical Association*, 91, 444–455.

Angrist, J. D. and J.-S. Pischke (2008): *Mostly harmless econometrics: An empiricist's companion*, Princeton: Princeton university press.

Ao, X., X. Chen, and Z. Zhao (2021): "Is Care by Grandparents or Parents Better for Children's Non-cognitive Skills? Evidence on Locus of Control from China," *IZA Discussion Paper*, 14183.

Armstrong, K., J. Fraser, M. Dadds, and J. Morris (1999): "A randomized, controlled trial of nurse home visiting to vulnerable families with newborns," *Journal of Paediatrics and Child Health*, 35, 237–244.

Arpino, B. and V. Bordone (2014): "Does grandparenting pay off? The effect of child care on grandparents' cognitive functioning," *Journal of Marriage and Family*, 76, 337–351.

——— (2017): "Regular provision of grandchild care and participation in social activities," *Review of Economics of the Household*, 15, 135–174.

Atalay, K. and G. F. Barrett (2014): “The causal effect of retirement on health: New evidence from Australian pension reform,” *Economics Letters*, 125, 392–395.

——— (2015): “The impact of age pension eligibility age on retirement and program dependence: Evidence from an Australian experiment,” *Review of Economics and Statistics*, 97, 71–87.

Atalay, K., G. F. Barrett, and A. Staneva (2019): “The effect of retirement on elderly cognitive functioning,” *Journal of Health Economics*, 66, 37–53.

Autorengruppe Bildungsberichterstattung (2020): *Bildung in Deutschland 2020. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt.*, Bielefeld: WBV Publikation.

Bach, M., J. Koebe, and F. Peter (2019): “Long Run Effects of Universal Childcare on Personality Traits,” *DIW Berlin Discussion Paper*, 1815.

Backhaus, A. and M. Barslund (2021): “The effect of grandchildren on grandparental labor supply: Evidence from Europe,” *European Economic Review*, 137, 103817.

Baker, M. (2011): “Innis Lecture: Universal early childhood interventions: what is the evidence base?” *Canadian Journal of Economics/Revue canadienne d’économique*, 44, 1069–1105.

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

——— (2019): “The long-run impacts of a universal child care program,” *American Economic Journal: Economic Policy*, 11, 1–26.

Baker, M. and K. Milligan (2008): “Maternal employment, breastfeeding, and health: Evidence from maternity leave mandates,” *Journal of Health Economics*, 27, 871–887.

Barlow, J., H. Davis, E. McIntosh, P. Jarrett, C. Mockford, and S. Stewart-Brown (2007): “Role of home visiting in improving parenting and health in families at risk of abuse and neglect: Results of a multicentre randomised controlled trial and economic evaluation,” *Archives of Disease in Childhood*, 92, 229–233.

Barnett, W. S. (1985): “Benefit-cost analysis of the Perry Preschool Program and its policy implications,” *Educational Evaluation and Policy Analysis*, 7, 333–342.

——— (2011): "Effectiveness of early educational intervention," *Science*, 333, 975–978.

Barnett, W. S. and L. N. Masse (2007): "Comparative benefit–cost analysis of the Abecedarian program and its policy implications," *Economics of Education Review*, 26, 113–125.

Baron, R., M. Taye, I. Besseling-van der Vaart, J. Ujčič-Voortman, H. Szajewska, J. C. Seidell, and A. Verhoeff (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, 335–349.

Barschkett, M., L. Gambaro, C. Schaeper, C. K. Spiess, and E. Ziege (2022a): "Oma und Opa gefragt? Veränderungen in der Enkelbetreuung - Wohlbefinden von Eltern - Wohlergehen von Kindern." Tech. rep., Bundesinstitut für Bevölkerungsforschung, DIW Berlin, Stiftung Ravensburger Verlag.

Barschkett, M., J. Geyer, P. Haan, and A. Hammerschmid (2022): "The effects of an increase in the retirement age on health—Evidence from administrative data," *The Journal of the Economics of Ageing*, 100403.

Barschkett, M., B. Koletzko, and C. K. Spiess (2021a): "COVID-19 Associated Contact Restrictions in Germany: Marked Decline in Children's Outpatient Visits for Infectious Diseases without Increasing Visits for Mental Health Disorders," *Children*, 8, 728.

Barschkett, M. and L. Schmitz (2020): "The cost-effectiveness in BRISE," Unpublished poster.

Barschkett, M., C. K. Spiess, and E. Ziege (2021b): "Does grandparenting pay off for the next generations? Intergenerational effects of grandparental care," *IZA Discussion Paper*, 14795.

Bauer, A. B. and R. Eichenberger (2021): "Worsening workers' health by lowering retirement age: The malign consequences of a benign reform," *The Journal of the Economics of Ageing*, 18, 100296.

Bauernschuster, S., T. Hener, and H. Rainer (2016): "Children of a (policy) revolution: The introduction of universal child care and its effect on fertility," *Journal of the European Economic Association*, 14, 975–1005.

Bauernschuster, S. and M. Schlotter (2015): "Public child care and mothers' labor supply—Evidence from two quasi-experiments," *Journal of Public Economics*, 123, 1–16.

Behncke, S. (2012): "Does retirement trigger ill health?" *Health Economics*, 21, 282–300.

Belloni, M., E. Meschi, and G. Pasini (2016): "The effect on mental health of retiring during the economic crisis," *Health Economics*, 25, 126–140.

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

Benson, J. B. and M. M. Haith (2010): *Language, memory, and cognition in infancy and early childhood*, Academic Press.

Berger, E. M. and C. K. Spiess (2011): "Maternal life satisfaction and child outcomes: Are they related?" *Journal of Economic Psychology*, 32, 142–158.

Berlinski, S. and S. Galiani (2007): "The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment," *Labour Economics*, 14, 665–680.

Besharov, D., P. Germanis, C. Higney, D. Call, and P. Germans (2011): "Assessments of twenty-six early childhood evaluations," Tech. rep., College Park, MD: University of Maryland, School of Public Policy.

Birch, L. L. (1999): "Development of food preferences," *Annual Review of Nutrition*, 19, 41–62.

Bíró, A. (2016): "Outpatient visits after retirement in Europe and the US," *International Journal of Health Economics and Management*, 16, 363–385.

Bíró, A. and P. Elek (2018): "How does retirement affect healthcare expenditures? Evidence from a change in the retirement age," *Health Economics*, 27, 803–818.

Blanden, J., E. Del Bono, K. Hansen, and B. Rabe (2022): "Quantity and quality of childcare and children's educational outcomes," *Journal of Population Economics*, 35, 785–828.

Bloemen, H., S. Hochguertel, and J. Zweerink (2017): "The causal effect of retirement on mortality: Evidence from targeted incentives to retire early," *Health Economics*, 26, e204–e218.

Bloom, D. E., D. Canning, and J. Sevilla (2004): "The effect of health on economic growth: A production function approach," *World development*, 32, 1–13.

Blossfeld, H.-P. and J. Von Maurice (2011): "Bildung als lebenslanger Prozess," *Zeitschrift für Erziehungswissenschaft*, 14, 19–34.

BMG (2016): "Glossar: Einheitlicher Bewertungsmaßstab - EBM," Tech. rep., Bundesgesundheitsministerium, `https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/e/einheitlicher-bewertungsmassstab-ebm.html`, accessed: 2022-12-12.

BMG (2020): *Das deutsche Gesundheitssystem*, `https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Gesundheit/Broschueren/200629_BMG_Das_deutsche_Gesundheitssystem_DE.pdf`: Bundesministerium für Gesundheit (BMG), accessed: 2022-12-13.

Boland, M. R., Z. Shahn, D. Madigan, G. Hripcsak, and N. P. Tatonetti (2015): "Birth month affects lifetime disease risk: A phenome-wide method," *Journal of the American Medical Informatics Association*, 22, 1042–1053.

Börsch-Supan, A. and C. B. Wilke (2004): "The German public pension system: How it was, how it will be," *NBER Discussion Paper*, w10525.

Borusyak, K., X. Jaravel, and J. Spiess (2021): "Revisiting event study designs: Robust and efficient estimation," *arXiv preprint*, 2108.12419.

Bosque-Mercader, L. (2022): "The Effect of a Universal Preschool Programme on Long-Term Health Outcomes: Evidence from Spain," *HEDG, c/o Department of Economics, University of York*, 22/06.

Bound, J., C. Brown, and N. Mathiowetz (2001): "Chapter 59 - Measurement Error in Survey Data," Elsevier, vol. 5 of *Handbook of Econometrics*, 3705–3843.

Bratsch-Hines, M. E., I. Mokrova, L. Vernon-Feagans, F. L. P. K. Investigators, et al. (2015): "Child care instability from 6 to 36 months and the social adjustment of children in prekindergarten," *Early Childhood Research Quarterly*, 30, 106–116.

Bratti, M., T. Frattini, and F. Scervini (2018): "Grandparental availability for child care and maternal labor force participation: Pension reform evidence from Italy," *Journal of Population Economics*, 31, 1239–1277.

Breivik, A.-L. (2020): “Determinants of Health and Labor Market Outcomes,” Ph.D. thesis, University of Bergen.

Bremberg, S. (2009): “A perfect 10: Why Sweden comes out on top in early child development programming,” *Paediatrics & child health*, 14, 677–680.

Briggs, N., J. Weatherhead, K. J. Sastry, and P. J. Hotez (2016): “The hygiene hypothesis and its inconvenient truths about helminth infections,” *PLoS Neglected Tropical Diseases*, 10, e0004944.

BRISE Consortium (2013): “Bremer Initiative zur Stärkung frühkindlicher Entwicklung (Bremen Initiative on Early Childhood Development). Project Proposal (unpublished),” .

——— (2022): “BRISE - Bremer Initiative zur Stärkung frühkindlicher Entwicklung,” Tech. rep., `https://www.brise-bremen.de/#top`.

Brockmann, H., R. Müller, and U. Helmert (2009): “Time to retire–Time to die? A prospective cohort study of the effects of early retirement on long-term survival,” *Social Science & Medicine*, 69, 160–164.

Brodeur, A. and M. Connolly (2013): “Do higher child care subsidies improve parental well-being? Evidence from Quebec’s family policies,” *Journal of Economic Behavior & Organization*, 93, 1–16.

Brüderl, J., R. Frister, K. Hajek, M. Herzig, R. Lenke, P. Schütze, and N. Schumann (2020): “Pairfam Data Manual. Release 11.0,” Tech. rep., LMU Munich.

Bubonya, M., D. A. Cobb-Clark, and M. Wooden (2017): “Mental health and productivity at work: Does what you do matter?” *Labour economics*, 46, 150–165.

Buda, S., K. Köpke, M. Luchtenberg, B. Schweiger, B. Biere, S. Duwe, L. Fiebig, U. Buchholz, W. Haas, et al. (2010): “Bericht zur Epidemiologie der Influenza in Deutschland Saison 2009/10,” Tech. rep., Robert Koch-Institut.

Burger, K. (2010): “How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds,” *Early childhood research quarterly*, 25, 140–165.

Callaway, B., A. Goodman-Bacon, and P. H. Sant’Anna (2021): “Difference-in-differences with a continuous treatment,” *arXiv preprint*, 2107.02637.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225, 200–230.

Camehl, G. F., C. K. Spiess, and K. Hahlweg (2020): "The effects of a parenting program on maternal well-being: Evidence from a Randomized Controlled Trial," *The BE Journal of Economic Analysis & Policy*, 20.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 90, 414–427.

Cameron, A. C. and P. K. Trivedi (2005): *Microeconometrics: Methods and applications*, Cambridge University Press.

Campbell, F. A. and C. T. Ramey (1991): "The Carolina Abecedarian Project," Tech. rep., ERIC.

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

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

——— (2016): "Partial insurance and investments in children," *The Economic Journal*, 126, F66–F95.

Carraro, S., N. Scheltema, L. Bont, and E. Baraldi (2014): "Early-life origins of chronic respiratory diseases: Understanding and promoting healthy ageing," *European Respiratory Journal*, 44, 1682–1696.

Cascio, E. U. (2009): "Maternal labor supply and the introduction of kindergartens into American public schools," *Journal of Human Resources*, 44, 140–170.

——— (2015): "The promises and pitfalls of universal early education," *IZA World of Labor*, 116.

Cattan, S., G. Conti, C. Farquharson, R. Ginja, and M. Pecher (2021): "The health effects of universal early childhood interventions: Evidence from Sure Start," *CEPR Discussion Paper*, DP16730.

CATTAN, S., D. KAMHÖFER, M. KARLSSON, AND T. NILSSON (2017): "The short- and long-term effects of student absence: Evidence from Sweden," *IZA Discussion Paper*, 10995.

CAWLEY, J. AND C. K. SPIESS (2008): "Obesity and skill attainment in early childhood," *Economics & Human Biology*, 6, 388–397.

CELIDONI, M. AND V. REBBA (2017): "Healthier lifestyles after retirement in Europe? Evidence from SHARE," *The European Journal of Health Economics*, 18, 805–830.

CHAPMAN, S. N., J. E. PETTAY, M. LAHDENPERÄ, AND V. LUMMAA (2018): "Grandmotherhood across the demographic transition," *PloS one*, 13, e0200963.

CHARLES, K. K. (2004): "Is retirement depressing? Labor force inactivity and psychological well-being in later life," *Research in Labor Economics*, 23, 269–299.

CHEN, Y. AND X. ZHANG (2018): "When mommies become nannies: The effects of parental retirement across generations," *SSRN*, 3146737.

CLARKE, D., J. P. ROMANO, AND M. WOLF (2020): "The Romano–Wolf multiple-hypothesis correction in Stata," *The Stata Journal*, 20, 812–843.

COE, N. AND G. ZAMARRO (2015): "Does retirement impact health care utilization?" *CESR-Schaeffer Working Paper*, 2015-032.

COE, N. B. AND G. ZAMARRO (2011): "Retirement effects on health in Europe," *Journal of Health Economics*, 30, 77–86.

COMPTON, J. AND R. A. POLLAK (2014): "Family proximity, childcare, and women's labor force attachment," *Journal of Urban Economics*, 79, 72–90.

CONEUS, K. AND C. K. SPIESS (2012a): "The intergenerational transmission of health in early childhood—Evidence from the German Socio-Economic Panel Study," *Economics & Human Biology*, 10, 89–97.

——— (2012b): "Pollution exposure and child health: Evidence for infants and toddlers in Germany," *Journal of Health Economics*, 31, 180–196.

CONTI, G., J. J. HECKMAN, AND R. PINTO (2016): "The effects of two influential early childhood interventions on health and healthy behaviour," *The Economic Journal*, 126, F28–F65.

Cools, S., J. H. Fiva, and L. J. Kirkebøen (2015): "Causal effects of paternity leave on children and parents," *The Scandinavian Journal of Economics*, 117, 801–828.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018): "Who benefits from universal child care? Estimating marginal returns to early child care attendance," *Journal of Political Economy*, 126, 2356–2409.

Cox, J. L., G. Chapman, D. Murray, and P. Jones (1996): "Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women," *Journal of Affective Disorders*, 39, 185–189.

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

Cunha, F. and J. J. Heckman (2008): "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation," *Journal of Human Resources*, 43, 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78, 883–931.

Currie, J. (2020): "Child health as human capital," *Health Economics*, 29, 452–463.

Currie, J. and D. Almond (2011): "Human capital development before age five," in *Handbook of labor economics*, Elsevier, vol. 4, 1315–1486.

Currie, J. and M. Stabile (2006): "Child mental health and human capital accumulation: The case of ADHD," *Journal of Health Economics*, 25, 1094–1118.

Dahlen, H. M. (2016): "The impact of maternal depression on child academic and socioemotional outcomes," *Economics of Education Review*, 52, 77–90.

Danielsbacka, M., A. O. Tanskanen, D. A. Coall, and M. Jokela (2019): "Grandparental childcare, health and well-being in Europe: A within-individual investigation of longitudinal data," *Social Science & Medicine*, 230, 194–203.

Danzer, N., M. Halla, N. Schneeweis, and M. Zweimüller (2022): "Parental leave,(in) formal childcare, and long-term child outcomes," *Journal of Human Resources*, 57, 1826–1884.

Datta Gupta, N. and M. Simonsen (2010): "Non-cognitive child outcomes and universal high quality child care," *Journal of Public Economics*, 94, 30–43.

Davis, K. and K. K. Christoffel (1994): "Obesity in preschool and school-age children: Treatment early and often may be best," *Archives of Pediatrics & Adolescent Medicine*, 148, 1257–1261.

Daysal, N. M., H. Ding, M. Rossin-Slater, and H. Schwandt (2022): "Germs in the Family: The Short-and Long-Term Consequences of Intra-Household Disease Spread," `http://www.meltemdaysal.com/docs/ChildSickness.pdf`, accessed: 2022-12-13.

De Chaisemartin, C. and X. d'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110, 2964–96.

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

Del Boca, D., M. Locatelli, and D. Vuri (2005): "Child-care choices by working mothers: The case of Italy," *Review of Economics of the Household*, 3, 453–477.

Del Boca, D., D. Piazzalunga, and C. Pronzato (2018): "The role of grandparenting in early childcare and child outcomes," *Review of Economics of the Household*, 16, 477–512.

Destatis (2018): "Statistiken der Kinder-und Jugendhilfe: Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderter Kindertagespflege am 01.03. 2018," Tech. rep., Statistisches Bundesamt (Destatis), `https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00096407`, accessed: 2022-12-13.

——— (2019a): "Kindertagesbetreuung regional," Tech. rep., Statistisches Bundesamt (Destatis), `https://www.statistikportal.de/de/karten/kindertagesbetreuung-regional`, accessed: 2022-12-13.

——— (2019b): "Statistiken der Kinder- und Jugendhilfe: Kinder und tätige Personen in Tageseinrichtungen in öffentlich geförderter Kindertagespflege am 01.03.2019," Tech. rep., Statistisches Bundesamt (Destatis), `https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00123476`, accessed: 2022-12-13.

DESTATIS (2020): "Statistischer Bericht - Budget für Bildung, Forschung und Wissenschaft 2019/2020," Tech. rep., Statistisches Bundesamt (Destatis), https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsfinanzen-Ausbildungsfoerderung/Publikationen/Downloads-Bildungsfinanzen/statistischer-bericht-budget-bildung-5217108207005.html?nn=209224, accessed: 2022-12-13.

DESTATIS (2021): "Elterngeld 2020: Väteranteil steigt auf knapp 25%," Tech. rep., Statistisches Bundesamt (Destatis), https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/03/PD21_146_22922.html, accessed: 2022-12-13.

——— (2022a): "Betreuungsquote von Kindern unter 6 Jahren nach Bundesländern," Tech. rep., Statistisches Bundesamt (Destatis), https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Soziales/Kindertagesbetreuung/Tabellen/betreuungsquote.html, accessed: 2022-12-13.

——— (2022b): "Erwerbstätigenquoten nach Gebietsstand und Geschlecht in der Altersgruppe 15 bis unter 65 Jahren Ergebnis des Mikrozensus in Prozent," Tech. rep., Statistisches Bundesamt (Destatis), https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbstaetigkeit/Tabellen/erwerbstaetigenquoten-gebietsstand-geschlecht-altergruppe-mikrozensus.html, accessed: 2022-12-13.

——— (2022c): "Statistiken der Kinder-und Jugendhilfe: Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderten Kindertagespflege am 01.03.2022," Tech. rep., Statistisches Bundesamt (Destatis), https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Soziales/Kindertagesbetreuung/Publikationen/_publikationen-innen-kinder-taetige-personen.html, accessed: 2022-12-13.

DIENER, E. AND M. Y. CHAN (2011): "Happy people live longer: Subjective well-being contributes to health and longevity," *Applied Psychology: Health and Well-Being*, 3, 1–43.

DIETRICHSON, J., I. LYKKE KRISTIANSEN, AND B. A. VIINHOLT (2020): "Universal preschool programs and long-term child outcomes: A systematic review," *Journal of Economic Surveys*, 34, 1007–1043.

DIMARIA, C. H., C. PERONI, AND F. SARRACINO (2020): "Happiness matters: Productivity gains from subjective well-being," *Journal of Happiness Studies*, 21, 139–160.

Doblhammer, G. and J. W. Vaupel (2001): "Lifespan depends on month of birth," *Proceedings of the National Academy of Sciences*, 98, 2934–2939.

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

Doyle, O., N. Fitzpatrick, J. Lovett, and C. Rawdon (2015): "Early intervention and child physical health: Evidence from a Dublin-based randomized controlled trial," *Economics & Human Biology*, 19, 224–245.

Drange, N. and T. Havnes (2019): "Early childcare and cognitive development: Evidence from an assignment lottery," *Journal of Labor Economics*, 37, 581–620.

DRV (2022): "Die richtige Altersrente für Sie," Tech. rep., Deutsche Rentenversicherung, `https://www.deutsche-rentenversicherung.de/SharedDocs/Downloads/DE/Broschueren/national/die_richtige_altersrente_fuer_sie.html`, accessed: 2022-12-13.

DuMont, K., K. Kirkland, S. Mitchell-Herzfeld, S. Ehrhard-Dietzel, M. L. Rodriguez, E. Lee, C. Layne, and R. Greene (2010): "A randomized trial of Healthy Families New York (HFNY): Does home visiting prevent child maltreatment," *Rensselaer, NY: New York State Office of Children & Family Services and Albany, NY: The University of Albany, State University of New York.*

Dunifon, R. E., C. E. Near, and K. M. Ziol-Guest (2018): "Backup parents, playmates, friends: Grandparents' time with grandchildren," *Journal of Marriage and Family*, 80, 752–767.

Eckenrode, J., M. I. Campa, P. A. Morris, C. R. Henderson Jr, K. E. Bolger, H. Kitzman, and D. L. Olds (2017): "The prevention of child maltreatment through the Nurse Family Partnership program: Mediating effects in a long-term follow-up study," *Child Maltreatment*, 22, 92–99.

Eibich, P. (2015): "Understanding the effect of retirement on health: Mechanisms and heterogeneity," *Journal of Health Economics*, 43, 1–12.

Eibich, P. and L. Goldzahl (2021): "Does retirement affect secondary preventive care use? Evidence from breast cancer screening," *Economics & Human Biology*, 43, 101061.

Eibich, P. and T. Siedler (2020): "Retirement, intergenerational time transfers, and fertility," *European Economic Review*, 124, 103392.

Eide, E. R. and M. H. Showalter (2011): "Estimating the relation between health and education: What do we know and what do we need to know?" *Economics of Education Review*, 30, 778–791.

Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2015): "Early childhood education," in *Economics of Means-Tested Transfer Programs in the United States, Volume 2*, University of Chicago Press, 235–297.

Enserink, R., R. Ypma, G. A. Donker, H. A. Smit, and W. van Pelt (2013): "Infectious disease burden related to child day care in the Netherlands," *The Pediatric Infectious Disease Journal*, 32, e334–e340.

Etgeton, S., B. Fischer, and H. Ye (2021): "The Effect of Increasing Retirement Age on Households' Savings and Consumption Expenditures," *University of Bonn and University of Mannheim Discussion Paper Series*, crctr224_2021_255v1.

Etgeton, S. and A. Hammerschmid (2019): "The Effect of Early Retirement on Health: Evidence from a German Pension Reform," *Mimeo*.

European Council (2002): "Presidency Conclusions Barcelona European Council 15 and 16 March 2002," Tech. rep., Barcelona European Council, `http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/71025.pdf`, accessed: 2022-12-13.

Eurostat (2012): "Active Ageing and Solidarity Between Generations — A Statistical Portrait of the European Union 2012," Tech. rep., European Commission, `https://ec.europa.eu/eurostat/documents/3217494/5740649/KS-EP-11-001-EN.PDF/1f0b25f8-3c86-4f40-9376-c737b54c5fcf`, accessed: 2022-12-13.

Federal Institute for Population Research (2020): "Erwerbstätigkeit von Eltern," Tech. rep., Federal Institute for Population Research, `https://www.demografie-portal.de/DE/Fakten/erwerbstaetigenquote-eltern.html`, accessed: 2022-12-13.

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

Felfe, C., N. Nollenberger, and N. Rodríguez-Planas (2015): "Can't buy mommy's love? Universal childcare and children's long-term cognitive development," *Journal of Population Economics*, 28, 393–422.

Fenoll, A. A. (2020): "The uneven impact of women's retirement on their daughters' employment," *Review of Economics of the Household*, 18, 795–821.

Fergusson, E., B. Maughan, and J. Golding (2008): "Which children receive grandparental care and what effect does it have?" *Journal of Child Psychology and Psychiatry*, 49, 161–169.

Fernald, A., V. A. Marchman, and A. Weisleder (2013): "SES differences in language processing skill and vocabulary are evident at 18 months," *Developmental Science*, 16, 234–248.

Fischer, B. and K.-U. Müller (2020): "Time to care? The effects of retirement on informal care provision," *Journal of Health Economics*, 73, 102350.

Fischer, M., M. Karlsson, and T. Nilsson (2013): "Effects of compulsory schooling on mortality: Evidence from Sweden," *International journal of environmental research and public health*, 10, 3596–3618.

Fitzpatrick, M. D. and T. J. Moore (2018): "The mortality effects of retirement: Evidence from Social Security eligibility at age 62," *Journal of Public Economics*, 157, 121–137.

Frijters, P., D. W. Johnston, and M. A. Shields (2014): "The effect of mental health on employment: Evidence from Australian panel data," *Health Economics*, 23, 1058–1071.

Frimmel, W., M. Halla, B. Schmidpeter, and R. Winter-Ebmer (2020): "Grandmothers' labor supply," *Journal of Human Resources*, 0419–10144R1.

Gambaro, L., M. Kreyenfeld, D. Schacht, and C. K. Spiess (2018): "Lebenszufriedenheit von Geflüchteten in Deutschland ist deutlich geringer, wenn ihre Kinder im Ausland leben," *DIW Wochenbericht*, 85, 905–916.

Garrouste, C. and E. Perdrix (2022): "Is there a consensus on the health consequences of retirement? A literature review," *Journal of Economic Surveys*, 36, 841–879.

Gelbach, J. B. (2002): "Public schooling for young children and maternal labor supply," *American Economic Review*, 92, 307–322.

Geyer, J., P. Haan, A. Hammerschmid, and M. Peters (2020): "Labor market and distributional effects of an increase in the retirement age," *Labour Economics*, 101–117.

GEYER, J. AND C. WELTEKE (2021): "Closing Routes to Retirement for Women How Do They Respond?" *Journal of Human Resources*, 56, 311–341.

GKV (2021): "Vergütung ärztlicher Leistungen," Tech. rep., GKV Spitzenverband, `https://www.gkv-spitzenverband.de/presse/themen/verguetung_aerztlicher_leistungen/s_thema_aerzteverguetung.jsp`, accessed: 2022-12-13.

GLAZNER, J., J. BONDY, D. LUCKEY, AND D. OLDS (2004): "Effect of the Nurse Family Partnership on Government Expenditures for Vulnerable First-Time Mothers and Their Children in Elmira, New York, Memphis, Tennessee, and Denver, Colorado (# 90XP0017): Final Report to the Administration for Children and Families," *Washington, DC: US Deptartment of Health and Human Services.*

GODARD, M. (2016): "Gaining weight through retirement? Results from the SHARE survey," *Journal of Health Economics*, 45, 27–46.

GODLEE, F. (2011): "What is health?" *British Medical Journal Publishing Group*, 343:d4817.

GOEBEL, J., M. M. GRABKA, S. LIEBIG, M. KROH, D. RICHTER, C. SCHRÖDER, AND J. SCHUPP (2019): "The german socio-economic panel (soep)," *Jahrbücher für Nationalökonomie und Statistik*, 239, 345–360.

GOHL, N., P. HAAN, E. KURZ, AND F. WEINHARDT (2020): "Working Life and Human Capital Investment: Causal Evidence from Pension Reform," *IZA Discussion Paper*, 12891.

GOODMAN, R. (1997): "The Strengths and Difficulties Questionnaire: A research note," *Journal of child psychology and psychiatry*, 38, 581–586.

GOODMAN-BACON, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225, 254–277.

GORRY, A., D. GORRY, AND S. N. SLAVOV (2018): "Does retirement improve health and life satisfaction?" *Health Economics*, 27, 2067–2086.

GRAY-LOBE, G., P. A. PATHAK, AND C. R. WALTERS (2021): "The long-term effects of universal preschool in Boston," *National Bureau of Economic Research Working Paper*, 28756.

Grip, A. d., M. Lindeboom, and R. Montizaan (2012): "Shattered dreams: The effects of changing the pension system late in the game," *The Economic Journal*, 122, 1–25.

Grosch, C. and G. Niebsch (1987): "Vergleich des Gesundheitszustandes von Kleinkindern, die in der Krippe bzw. ausschließlich in der Familie erzogen werden," *Hygiene des Kindes-und Jugendalters*, 10, 78–82.

Grossman, M. (1972): "On the concept of health capital and the demand for health," *Journal of Political economy*, 80, 223–255.

Grøtting, M. W. and O. S. Lillebø (2020): "Health effects of retirement: Evidence from survey and register data," *Journal of Population Economics*, 33, 671–704.

Haeck, C., L. Lebihan, and P. Merrigan (2018): "Universal child care and long-term effects on child well-being: Evidence from Canada," *Journal of Human Capital*, 12, 38–98.

Hagen, J. (2018): "The effects of increasing the normal retirement age on health care utilization and mortality," *Journal of Population Economics*, 31, 193–234.

Hainmueller, J. (2012): "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies," *Political Analysis*, 20, 25–46.

Hallberg, D., P. Johansson, and M. Josephson (2015): "Is an early retirement offer good for your health? Quasi-experimental evidence from the army," *Journal of Health Economics*, 44, 274–285.

Hank, K. and I. Buber (2009): "Grandparents caring for their grandchildren: Findings from the 2004 Survey of Health, Ageing, and Retirement in Europe," *Journal of family Issues*, 30, 53–73.

Hank, K., G. Cavrini, G. Di Gessa, and C. Tomassini (2018): "What do we know about grandparents? Insights from current quantitative data and identification of future data needs," *European Journal of Ageing*, 15, 225–235.

Hank, K., M. Kreyenfeld, and C. K. Spiess (2004): "Kinderbetreuung und Fertilität in Deutschland/Child Care and Fertility in Germany," *Zeitschrift für Soziologie*, 33, 228–244.

Havnes, T. and M. Mogstad (2011a): "Money for nothing? Universal child care and maternal employment," *Journal of Public Economics*, 95, 1455–1465.

——— (2011b): "No child left behind: Subsidized child care and children's long-run outcomes," *American Economic Journal: Economic Policy*, 3, 97–129.

——— (2015): "Is universal child care leveling the playing field?" *Journal of Public Economics*, 127, 100–114.

Hawkins, S. S., S. Hristakeva, M. Gottlieb, and C. F. Baum (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., R. Pinto, and P. Savelyev (2013): "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review*, 103, 2052–86.

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

Heckman, J. J., M. L. Holland, K. K. Makino, R. Pinto, and M. Rosales-Rueda (2017): "An analysis of the memphis nurse-family partnership program," *National Bureau of Economic Research Working Paper*, w23610.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010): "The rate of return to the HighScope Perry Preschool Program," *Journal of Public Economics*, 94, 114–128.

Heckman, J. J. and S. Mosso (2014): "The economics of human development and social mobility," *National Bureau of Economic Research Working Paper*, w19925.

Heller-Sahlgren, G. (2017): "Retirement blues," *Journal of Health Economics*, 54, 66–78.

Hellevik, O. (2009): "Linear versus logistic regression when the dependent variable is a dichotomy," *Quality & Quantity*, 43, 59–74.

Herbst, C. M. and E. Tekin (2014): "Child care subsidies, maternal health, and child–parent interactions: Evidence from three nationally representative datasets," *Health Economics*, 23, 894–916.

Hernaes, E., S. Markussen, J. Piggott, and O. L. Vestad (2013): "Does retirement age impact mortality?" *Journal of Health Economics*, 32, 586–598.

Hetschko, C., A. Knabe, and R. Schöb (2014): "Changing identity: Retiring from unemployment," *The Economic Journal*, 124, 149–166.

Hong, K., K. Dragan, and S. Glied (2019): "Seeing and hearing: The impacts of New York City's universal pre-kindergarten program on the health of low-income children," *Journal of Health Economics*, 64, 93–107.

Horner, E. M. and M. R. Cullen (2016): "The impact of retirement on health: Quasi-experimental methods using administrative data," *BMC Health Services Research*, 16.

Huber, M., J. A. Knottnerus, L. Green, H. Van Der Horst, A. R. Jadad, D. Kromhout, B. Leonard, K. Lorig, M. I. Loureiro, J. W. Van der Meer, et al. (2011): "How should we define health?" *Bmj*, 343.

Huebener, M. (2022): "The effects of education on health: An intergenerational perspective," *Journal of Human Resources*, 0219–10060R2.

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

Huinink, J., J. Brüderl, B. Nauck, S. Walper, L. Castiglioni, and M. Feldhaus (2011): "Panel analysis of intimate relationships and family dynamics (pairfam): Conceptual framework and design," *ZfF–Zeitschrift für Familienforschung/Journal of Family Research*, 23.

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

IAW (2018): "Armutspolitik in Bremen," Tech. rep., Institut Arbeit und Wirtschaft, `https://www.soziales.bremen.de/soziales/armuts-und-reichtumsberichte-70849`, accessed: 2022-12-13.

Jessen, J. (2022): "Culture, children and couple gender inequality," *European Economic Review*, 150, 104310.

Jessen, J., S. Schmitz, and S. Waights (2020): "Understanding day care enrolment gaps," *Journal of Public Economics*, 190, 104252.

Jungmann, T., T. Brand, V. Dähne, P. Herrmann, H. Günay, M. Sandner, and S. Sierau (2015): "Comprehensive evaluation of the Pro Kind home visiting program: A summary of results," *Mental Health & Prevention*, 3, 89–97.

Kahn, H. S., T. M. Morgan, L. D. Case, D. Dabelea, E. J. Mayer-Davis, J. M. Lawrence, S. M. Marcovina, G. Imperatore, S. for Diabetes in Youth Study Group, et al. (2009): "Association of type 1 diabetes with month of birth among US youth: The SEARCH for Diabetes in Youth Study," *Diabetes care*, 32, 2010–2015.

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

Kämpfen, F. and J. Maurer (2016): "Time to burn (calories)? The impact of retirement on physical activity among mature Americans," *Journal of Health Economics*, 45, 91–102.

Kamtsiuris, P., E. Bergmann, P. Rattay, and M. Schlaud (2007): "Inanspruchnahme medizinischer Leistungen: Ergebnisse des Kinder- und Jugendgesundheitssurveys (KiGGS)," in *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung, vol. 50.

Kanji, S. (2018): "Grandparent care: A key factor in mothers' labour force participation in the UK," *Journal of Social Policy*, 47, 523–542.

Karoly, L. A. (2012): "Toward standardization of benefit-cost analysis of early childhood interventions," *Journal of Benefit-Cost Analysis*, 3, 1–45.

——— (2017): "Investing in the Early Years," Tech. rep., Rand Cooperation.

Karoly, L. A., P. W. Greenwood, S. S. Everingham, J. Houbé, and M. R. Kilburn (1998): *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions*, Rand Corporation.

Karoly, L. A., M. R. Kilburn, and J. S. Cannon (2006): *Early childhood interventions: Proven results, future promise*, Rand Corporation.

Kassenärztliche Bundesvereinigung KBV (2019): "Honorarbericht," Tech. rep., Kassenärztliche Bundesvereinigung, `https://www.kbv.de/html/honorarbericht.php`, accessed: 2022-12-13.

——— (2021a): "Honorarberechnung in 300 Sekunden," Tech. rep., Kassenärztliche Bundesvereinigung, `https://www.kbv.de/html/10994.php`, accessed: 2022-12-13.

——— (2021b): "Honorarverhandlungen 2021: Schiedsspruch gegen Ärzteseite," Tech. rep., Kassenärztliche Bundesvereinigung, `https://www.kbv.de/html/2054.php`, accessed: 2022-12-13.

——— (2021c): "Honorarverteilung und -berechnung," Tech. rep., Kassenärztliche Bundesvereinigung, `https://www.kbv.de/html/1019.php`, accessed: 2022-12-13.

Kautz, T., J. Heckman, R. Diris, B. Ter Weel, and L. Borghans (2014): "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success," *National Bureau of Economic Research Working Paper*, w20749.

Kelly, Y. and M. Bartley (2010): "Parental and child health," *Children of the 21st century: the first five years*, 2, 249–272.

Kemptner, D., H. Jürges, and S. Reinhold (2011): "Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany," *Journal of Health Economics*, 30, 340–354.

Kim, J. H., W. Schulz, T. Zimmermann, and K. Hahlweg (2018): "Parent–child interactions and child outcomes: Evidence from randomized intervention," *Labour Economics*, 54, 152–171.

Klebanov, P. K., J. Brooks-Gunn, and M. C. McCormick (2001): "Maternal coping strategies and emotional distress: Results of an early intervention program for low birth weight young children." *Developmental Psychology*, 37, 654.

Korpi, W. and J. Palme (1998): "The paradox of redistribution and strategies of equality: Welfare state institutions, inequality, and poverty in the Western countries," *American sociological review*, 661–687.

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

Kröll, A. and R. Borck (2013): "The influence of child care on maternal health and mother-child interaction," *CESifo Working Paper Series*, 4289.

Kuehnle, D. and M. Oberfichtner (2020): "Does starting universal childcare earlier influence children's skill development?" *Demography*, 57, 61–98.

Kuger, S., J. Marcus, and C. K. Spiess (2019): "Day care quality and changes in the home learning environment of children," *Education Economics*, 27, 265–286.

KUHN, D., Y. BEN-SHLOMO, J. LYNCH, J. HALLQVIST, AND C. POWER (2003): "Life course epidemiology," *Journal of epidemiology and community health*, 57, 778.

KUHN, A., S. STAUBLI, J.-P. WUELLRICH, AND J. ZWEIMÜLLER (2019): "Fatal attraction? Extended unemployment benefits, labor force exits, and mortality," *Journal of Public Economics*, 191, 104087.

KUUSI, T., P. MARTIKAINEN, AND T. VALKONEN (2020): "The influence of old-age retirement on health: Causal evidence from the Finnish register data," *The Journal of the Economics of Ageing*, 100257.

LAUBER, V. (2015): "Essays in Empirical Family Economics: Investments in Children and Work-Life Balance of Parents," Ph.D. thesis, Freie Universität Berlin.

LEIMER, B. AND R. VAN EWIJK (2022): "No "Honeymoon Phase": Whose health benefits from retirement and when," *Economics & Human Biology*, 47, 101171.

LEVIN, H. M. AND P. J. MCEWAN (2000): *Cost-effectiveness analysis: Methods and applications*, vol. 4, Sage.

LIKERT, R. (1932): "A technique for the measurement of attitudes." *Archives of Psychology*, 22140.

LINDSAY, G., S. STRAND, AND H. DAVIS (2011): "A comparison of the effectiveness of three parenting programmes in improving parenting skills, parent mental-well being and children's behaviour when implemented on a large scale in community settings in 18 English local authorities: the parenting early intervention pathfinder (PEIP)," *BMC Public Health*, 11, 1–13.

LOVE, J. M., E. E. KISKER, C. ROSS, H. RAIKES, J. CONSTANTINE, K. BOLLER, J. BROOKS-GUNN, R. CHAZAN-COHEN, L. B. TARULLO, C. BRADY-SMITH, ET AL. (2005): "The effectiveness of Early Head Start for 3-year-old children and their parents: lessons for policy and programs." *Developmental Psychology*, 41, 885.

LOWENSTEIN, A. AND V. BENGTSON (2003): "Challenges of global aging to families in the twenty-first century," *Global aging and challenges to families*, 371–379.

LUCIFORA, C. AND D. VIGANI (2018): "Health care utilization at retirement: The role of the opportunity cost of time," *Health Economics*, 27, 2030–2050.

LUNDIN, D., E. MÖRK, AND B. ÖCKERT (2008): "How far can reduced childcare prices push female labour supply?" *Labour Economics*, 15, 647–659.

MacMillan, H. L., C. N. Wathen, J. Barlow, D. M. Fergusson, J. M. Leventhal, and H. N. Taussig (2009): "Interventions to prevent child maltreatment and associated impairment," *The Lancet*, 373, 250–266.

Maier-Pfeiffer, A., K. Kutz, and M. Sandner (2013): "Kosten, Nutzen und Finanzierung in den Frühen Hilfen," *Kinder schützen, Familien stärken*, 111–124.

Mazzonna, F. and F. Peracchi (2017): "Unhealthy retirement?" *Journal of Human Resources*, 52, 128–151.

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

McIntosh, E., J. Barlow, H. Davis, and S. Stewart-Brown (2009): "Economic evaluation of an intensive home visiting programme for vulnerable families: A cost-effectiveness analysis of a public health intervention," *Journal of Public Health*, 31, 423–433.

Miller, T. R. and D. Hendrie (2015): "Nurse Family Partnership: Comparing costs per family in randomized trials versus scale-up," *The Journal of Primary Prevention*, 36, 419–425.

Milovanska-Farrington, S. (2021): "The Effect of Parental and Grandparental Supervision Time Investment on Children's Early-Age Development," *Research in Economics*, 75, 286–304.

Motegi, H., Y. Nishimura, and K. Terada (2016): "Does retirement change lifestyle habits?" *The Japanese Economic Review*, 67, 169–191.

Muehler, G. (2010): "Consequences of mixed provision of child care: An overview on the German market," *ZEW Discussion Papers*, 08-077.

Müller, K.-U. and K. Wrohlich (2016): "Two steps forward—one step back? Evaluating contradicting child care policies in Germany," *CESifo Economic Studies*, 62, 672–698.

——— (2020): "Does Subsidized Care for Toddlers Increase Maternal Labor Supply?: Evidence from a Large-Scale Expansion of Early Childcare," *Labour Economics*, 62.

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

Neumann, K., P. Gierling, M. Albrecht, J. Dietzel, and A. Wolfschütz (2014): *Reform der ärztlichen Vergütung im ambulanten Sektor - Studienbericht für die Techniker Krankenkasse*, IGES Institut GmbH.

Nielsen, N. F. (2019): "Sick of retirement?" *Journal of Health Economics*, 65, 133–152.

Nishimura, Y., M. Oikawa, and H. Motegi (2018): "What explains the difference in the effect of retirement on health? Evidence from global aging data," *Journal of Economic Surveys*, 32, 792–847.

Nollenberger, N. and N. Rodríguez-Planas (2015): "Full-time universal childcare in a context of low maternal employment: Quasi-experimental evidence from Spain," *Labour Economics*, 36, 124–136.

OECD (2019a): "Education at a Glance 2019," Tech. rep., OECD, `https://www.oecd-ilibrary.org/content/publication/f8d7880d-en`, accessed: 2022-12-13.

——— (2019b): "Germany: Country Health Profile 2019, State of Health in the EU," Tech. rep., OECD, `https://www.oecd.org/publications/germany-country-health-profile-2019-36e21650-en.htm`, accessed: 2022-12-13.

OECD (2020): "OECD family database," Tech. rep., OECD, `http://www.oecd.org/social/family/database.htm`, accessed: 2022-12-13.

OECD (2021): "Education at a Glance 2021: OECD Indicators on Early Childhood Education and Care," Tech. rep., OECD, `https://www.oecd.org/education/education-at-a-glance/`, accessed: 2022-12-13.

——— (2022a): "Employment: Labour force participation rate, by sex and age group," Tech. rep., OECD.Stat, `https://stats.oecd.org/index.aspx?queryid=103872`, accessed: 2022-12-13.

——— (2022b): "OECD Health Statistics 2022 ," Tech. rep., OECD, `https://www.oecd.org/health/health-data.htm`, accessed: 2022-12-13.

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

Olds, D. L. (2006): "The nurse–family partnership: An evidence-based preventive intervention," *Infant Mental Health Journal*, 27, 5–25.

Olds, D. L., J. Robinson, R. O'Brien, D. W. Luckey, L. M. Pettitt, C. R. Henderson Jr, R. K. Ng, K. L. Sheff, J. Korfmacher, S. Hiatt, et al. (2002): "Home visiting by paraprofessionals and by nurses: A randomized, controlled trial," *Pediatrics*, 110, 486–496.

Oswald, A. J., E. Proto, and D. Sgroi (2015): "Happiness and productivity," *Journal of Labor Economics*, 33, 789–822.

Pauen, S. (2011): "Vom Baby zum Kleinkind," in *Vom Baby zum Kleinkind*, Springer, 1–11.

Pauen, S., L. Heilig, D. Danner, J. Haffner, A. Tettenborn, and J. Roos (2012): "Milestones of Normal Development in Early Years (MONDEY)," *Frühe Bildung*.

Peacock, S., S. Konrad, E. Watson, D. Nickel, and N. Muhajarine (2013): "Effectiveness of home visiting programs on child outcomes: A systematic review," *BMC public health*, 13, 1–14.

Pedron, S., W. Maier, A. Peters, B. Linkohr, C. Meisinger, W. Rathmann, P. Eibich, and L. Schwettmann (2020): "The effect of retirement on biomedical and behavioral risk factors for cardiovascular and metabolic disease," *Economics & Human Biology*, 38, 100893.

Peet, E. D., D. C. McCoy, G. Danaei, M. Ezzati, W. Fawzi, M.-R. Jarvelin, D. Pillas, and G. Fink (2015): "Early childhood development and schooling attainment: Longitudinal evidence from British, Finnish and Philippine birth cohorts," *PloS one*, 10, e0137219.

Pelkowski, J. M. and M. C. Berger (2004): "The impact of health on employment, wages, and hours worked over the life cycle," *The Quarterly Review of Economics and Finance*, 44, 102–121.

Perdrix, E. (2021): "Does later retirement change your healthcare consumption? Evidence from France," *Hal SHS Discussion Paper*, 02904339.

Peter, F. (2013): "Essays on children's non-cognitive skills and health," Ph.D. thesis, Freie Universität Berlin.

Peter, F. H., P. S. Schober, and K. C. Spiess (2016): "Early birds in day care: The social gradient in starting day care and children's non-cognitive skills," *CESifo Economic Studies*, 62, 725–751.

Phillips, D. and J. B. Young (2000): "Birth weight, climate at birth and the risk of obesity in adult life," *International journal of obesity*, 24, 281–287.

Pilipiec, P., W. Groot, and M. Pavlova (2020): "The Effect of an Increase of the Retirement Age on the Health, Well-Being, and Labor Force Participation of Older Workers: A Systematic Literature Review," *Journal of Population Ageing*, 1–45.

Polanska, K., J. Jurewicz, and W. Hanke (2015): "Smoking and alcohol drinking during pregnancy as the risk factors for poor child neurodevelopment – A review of epidemiological studies," *International Journal of Occupational Medicine and Environmental Health*, 28, 419.

Pönkä, A., T. Nurmi, E. Salminen, and E. Nykyri (1991): "Infections and other illnesses of children in day-care centers in Helsinki I: Incidences and effects of home and day-care center variables," *Infection*, 19, 230–236.

Pötzsch, O. (2012): "Geburtenfolge und Geburtenabstand – neue Daten und Befunde," *Wirtschaft und Statistik*, 2, 89–101.

Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn, and J. Downer (2012): "Third Grade Follow-Up to the Head Start Impact Study: Final Report. OPRE Report 2012-45." *Administration for Children & Families.*

Rainer, H., S. Bauernschuster, W. Auer, N. Danzer, M. Hancioglu, B. Hartmann, T. Hener, C. Holzner, N. Ott, J. Reinkowski, et al. (2013): *Kinderbetreuung*, 59, ifo Forschungsberichte.

Rindfuss, R. R., D. K. Guilkey, S. P. Morgan, and Ø. Kravdal (2010): "Child-care availability and fertility in Norway," *Population and Development Review*, 36, 725–748.

RKI (2010a): "Bericht zur Epidemiologie der Influenza in Deutschland Saison 2009/10," Tech. rep., Robert-Koch-Institut, Arbeitgemeinschaft Influenza, `https://edoc.rki.de/handle/176904/3377`, accessed: 2022-12-13.

——— (2010b): "Faktenblätter – Inanspruchnahme von Leistungen des Gesundheitssystems: Arztbesuch," Tech. rep., Robert-Koch-Institut, Epidemiologie und Gesundheitsberichterstattung, `https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/Faktenblaetter/GEDA10/geda10_fb_inhalt.html`, accessed: 2022-12-13.

RKI (2022): "Survstat RKI 2.0," Tech. rep., Robert-Koch-Institut, `https://survstat.rki.de/Content/Query/Create.aspx`, accessed: 2022-12-13.

Rohwedder, S. and R. J. Willis (2010): "Mental retirement," *Journal of Economic Perspectives*, 24, 119–138.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019): "Fast and wild: Bootstrap inference in Stata using boottest," *The Stata Journal*, 19, 4–60.

Rose, L. (2020): "Retirement and health: Evidence from England," *Journal of Health Economics*, 73, 102352.

Rupert, P. and G. Zanella (2018): "Grandchildren and their grandparents' labor supply," *Journal of Public Economics*, 159, 89–103.

Rustichini, A., W. G. Iacono, and M. McGue (2017): "The contribution of skills and family background to educational mobility," *The Scandinavian Journal of Economics*, 119, 148–177.

Sadler, L. S., A. Slade, N. Close, D. L. Webb, T. Simpson, K. Fennie, and L. C. Mayes (2013): "Minding the baby: Enhancing reflectiveness to improve early health and relationship outcomes in an interdisciplinary home-visiting program," *Infant Mental Health Journal*, 34, 391–405.

Sadruddin, A. F., L. A. Ponguta, A. L. Zonderman, K. S. Wiley, A. Grimshaw, and C. Panter-Brick (2019): "How do grandparents influence child health and development? A systematic review," *Social Science & Medicine*, 239, 112476.

Sammons, P., J. Hall, R. Smees, J. Goff, K. Sylva, T. Smith, M. Evangelou, N. Eisenstadt, and G. Smith (2015): *The impact of children's centres: studying the effects of children's centres in promoting better outcomes for young children and their families: Evaluation of Children's Centres in England (ECCE, Strand 4)*, Department for Education.

Sandner, M. (2013): "Effects of early childhood intervention on child development and early skill formation: Evidence from a randomized controlled trail," *Hannover Economic Papers (HEP)*, 518.

——— (2019): "Effects of early childhood intervention on fertility and maternal employment: Evidence from a randomized controlled trial," *Journal of Health Economics*, 63, 159–181.

Sandner, M., T. Cornelissen, T. Jungmann, and P. Herrmann (2018): "Evaluating the effects of a targeted home visiting program on maternal and child health outcomes," *Journal of Health Economics*, 58, 269–283.

Sandner, M. and T. Jungmann (2017): "Gender-specific effects of early childhood intervention: Evidence from a randomized controlled trial," *Labour Economics*, 45, 59–78.

Sandner, M. and S. L. Thomsen (2020): "Preventing child maltreatment: Beneficial side effects of public childcare provision," *Hannover Economic Papers (HEP)*, 669.

Sann, A. and K. Thrum (2005): "Opstapje–Schritt für Schritt," *Abschlussbericht des Modellprojekts, München: DJI.*

Sartorius, N. (2006): "The meanings of health and its promotion," *Croatian medical journal*, 47, 662.

Schaub, B., R. Lauener, and E. von Mutius (2006): "The many faces of the hygiene hypothesis," *Journal of Allergy and Clinical Immunology*, 117, 969–977.

Schaub, S., E. Ramseier, A. Neuhauser, S. C. Burkhardt, and A. Lanfranchi (2019): "Effects of home-based early intervention on child outcomes: A randomized controlled trial of Parents as Teachers in Switzerland," *Early Childhood Research Quarterly*, 48, 173–185.

Schienkiewitz, A., A.-K. Brettschneider, S. Damerow, and A. Schaffrath Rosario (2018): "Overweight and obesity among children and adolescents in Germany. Results of the cross-sectional KiGGS Wave 2 study and trends," Tech. rep., Robert Koch-Institut.

Schmitz, S. (2020): "The impact of publicly funded childcare on parental well-being: Evidence from cut-off rules," *European Journal of Population*, 36, 171–196.

Schmitz, S. and A. Kröger (2017): "Effizienzanalysen frühkindlicher Bildungs-und Betreuungsangebote: Mehr differenzierte Analysen für Deutschland erforderlich," Tech. rep., DIW Roundup: Politik im Fokus.

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

Schober, P. and C. Schmitt (2017): "Day-care availability, maternal employment and satisfaction of parents: Evidence from cultural and policy variations in Germany," *Journal of European Social Policy*, 27, 433–446.

Schober, P. S. and J. F. Stahl (2016): "Expansion of full-day childcare and subjective well-being of mothers: Interdependencies with culture and resources," *European Sociological Review*, 32, 593–606.

Schütte, K., O. Köller, Y. Anders, and F. Petermann (2020): "Empirische Arbeit: Kontinuität als Schlüssel zu wirksamer Förderung frühkindlicher Entwicklung benachteiligter Kinder," *Psychologie in Erziehung und Unterricht.*

Schönberg, U. and J. Ludsteck (2014): "Expansions in maternity leave coverage and mothers' labor market outcomes after childbirth," *Journal of Labor Economics*, 32, 469–505.

Shai, O. (2018): "Is retirement good for men's health? Evidence using a change in the retirement age in Israel," *Journal of Health Economics*, 57, 15–30.

Sidor, A., E. Kunz, A. Eickhorst, and M. Cierpka (2013): "Effects of the early prevention program "Keiner faellt durchs Netz"("Nobody slips through the net") on child, mother, and their relationship: A controlled study," *Infant Mental Health Journal*, 34, 11–24.

Siedentop, S., R. Junesch, M. Klein, R. Krumm, and R. Kleimann (2014): "Wanderungsmotive im ländlichen Raum," *Ministerium für Ländlichen Raum und Verbraucherschutz Baden-Württemberg, Stuttgart.*

Siegle, D. (2020): "Likert Scales," Tech. rep., `https://researchbasics.education.uconn.edu/likert_scales/`, accessed: 2022-12-13.

Silles, M. A. (2009): "The causal effect of education on health: Evidence from the United Kingdom," *Economics of Education review*, 28, 122–128.

Silvestrin, S., V. N. Hirakata, C. H. da Silva, and M. Z. Goldani (2020): "Inequalities in birth weight and maternal education: A time-series study from 1996 to 2013 in Brazil," *Scientific reports*, 10, 1–8.

Song, X. and R. D. Mare (2019): "Shared lifetimes, multigenerational exposure, and educational mobility," *Demography*, 56, 891–916.

Spiess, C. K. (2011): "Vereinbarkeit von Familie und Beruf–wie wirksam sind deutsche „Care Policies"?" *Perspektiven der Wirtschaftspolitik*, 12, 4–27.

——— (2013): "Effizienzanalysen frühkindlicher Bildungs-und Betreuungsprogramme," *Zeitschrift für Erziehungswissenschaft*, 16, 333–354.

Spiess, C. K. (2022): "Kita-Ökonomik–eine Perspektive für Deutschland," *Perspektiven der Wirtschaftspolitik*, 23, 25–37.

St. Pierre, R. G. and J. I. Layzer (1999): "Using home visits for multiple purposes: the Comprehensive Child Development Program," *The Future of children*, 134–151.

Stanat, P., S. Schipolowski, N. Mahler, S. Weirich, and S. Henschel (2019): *IQB-Bildungstrend 2018*, Münster; New York: Waxmann.

Stanat, P., S. Schipolowski, C. Rjosk, S. Weirich, and N. Haag (2017): *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*, Münster; New York: Waxmann.

Statista (2021): "Anzahl der Geburten in Deutschland von 1991 bis 2020," Tech. rep., `https://de.statista.com/statistik/daten/studie/235/umfrage/anzahl-der-geburten-seit-1993/`, accessed: 2022-12-13.

Staubli, S. and J. Zweimüller (2013): "Does raising the early retirement age increase employment of older workers?" *Journal of Public Economics*, 108, 17–32.

Stewart, W. F., J. A. Ricci, E. Chee, S. R. Hahn, and D. Morganstein (2003): "Cost of lost productive work time among US workers with depression," *Jama*, 289, 3135–3144.

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

Sweet, M. A. and M. I. Appelbaum (2004): "Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children," *Child development*, 75, 1435–1456.

Thompson, M., T. A. Vodicka, P. S. Blair, D. I. Buckley, C. Heneghan, and A. D. Hay (2013): "Duration of symptoms of respiratory tract infections in children: Systematic review," *Bmj*, 347, f7027.

UN (2022): "Transforming our World: The 2030 Agenda for Sustainable Development," Tech. rep., United Nations, `https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981`, accessed: 2022-12-13.

U.S. Department of Health & Human Services (2004): "Head Start Program Facts: Fiscal Year 2004," Tech. rep., `https://eclkc.ohs.acf.hhs.gov/archive/about-us/article/head-start-program-facts-fiscal-year-2004`, accessed: 2022-12-13.

Vaiserman, A. and M. Khalangot (2008): "Similar seasonality of birth in type 1 and type 2 diabetes patients: a sign for common etiology?" *Medical hypotheses*, 71, 604–605.

van den Berg, G. J. and B. M. Siflinger (2022): "The effects of a daycare reform on health in childhood–Evidence from Sweden," *Journal of Health Economics*, 81, 102577.

van der Heide, I., R. M. van Rijn, S. J. Robroek, A. Burdorf, and K. I. Proper (2013): "Is retirement good for your health? A systematic review of longitudinal studies," *BMC public health*, 13, 1–11.

van Huizen, T. and J. Plantenga (2018): "Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments," *Economics of Education Review*, 66, 206–222.

van Ours, J. C. and M. Picchio (2020): "Mental Health Effects of Retirement," *De Economist*, 168, 419–452.

Wager, S. and S. Athey (2018): "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 113, 1228–1242.

Wagner, M. M. and S. L. Clayton (1999): "The Parents as Teachers program: Results from two demonstrations," *The Future of Children*, 91–115.

Walendzik, A. and J. Wasem (2019): *Vergütung ambulanter und ambulant erbringbarer Leistungen*, Bertelsmann Stiftung.

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

Waters, E., A. de Silva-Sanigorski, B. J. Burford, T. Brown, K. J. Campbell, Y. Gao, R. Armstrong, L. Prosser, and C. D. Summerbell (2011): "Interventions for preventing obesity in children," *Cochrane database of systematic reviews*, 12.

WATTIE, N., C. I. ARDERN, AND J. BAKER (2008): "Season of birth and prevalence of overweight and obesity in Canada," *Early human development*, 84, 539–547.

WELL, D. N. (2007): "Accounting for the effect of health on economic growth," *The Quarterly Journal of Economics*, 122, 1265–1306.

WETZEL, M., N. SCHUMANN, AND C. SCHMIEDEBERG (2021): "New weights for the pairfam anchor data - Pairfam Technical Paper," Tech. rep., https://www.pairfam.de/fileadmin/user_upload/redakteur/publis/technical_papers/pairfam_Technical_Paper_17.pdf, accessed: 2022-12-13.

WHEELOCK, J. AND K. JONES (2002): "Grandparents are the next best thing: informal childcare for working parents in urban Britain," *Journal of social policy*, 31, 441–463.

WHO (1948): "Official Records of the World Health Organization No. 2: Summary Report on Proceedings, Minutes and Final Acts of the International Health Conference Held in New York From 19 June to 22 July 1946," Tech. rep., United Nations, World Health Organization, Interim Commission, https://apps.who.int/iris/handle/10665/85573, accessed: 2022-12-13.

WORKMAN, S. (2018): "Where does your child care dollar go? Understanding the true cost of quality early childhood education," *Washington: Center for American Progress*.

WROHLICH, K. (2008): "The excess demand for subsidized child care in Germany," *Applied Economics*, 40, 1217–1228.

ZAMBERLETTI, J., G. CAVRINI, AND C. TOMASSINI (2018): "Grandparents providing childcare in Italy," *European journal of ageing*, 15, 265–275.

ZHANG, Y., M. SALM, AND A. VAN SOEST (2018): "The effect of retirement on healthcare utilization: Evidence from China," *Journal of Health Economics*, 62, 165–177.

ZI (2015): "Die 50 häufigsten ICD-10-Schlüsselnummern nach Fachgruppen," Tech. rep., Zentralinstitut für die Kassenärztliche Versorgung in Deutschland, Ansprechpartner: Joachim Heuer, https://www.zi.de/fileadmin/images/content/PDFs_alle/Die_50_h%C3%A4ufigsten_ICD-2015.pdf.