Essays in Applied Microeconometrics

Inaugural-Dissertation
an der
Ludwig-Maximilians-Universität München

von
Marc Fabel

Essays in Applied Microeconometrics

Inauguraldissertation
zur Erlangung des akademischen Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Volkswirtschaftlichen Fakultät
der Ludwig-Maximilians-Universität München

2021

vorgelegt von
Marc G. Fabel

Referent:	Prof. Helmut Rainer, Ph.D.
Korreferent:	Prof. Dr. Stefan Bauernschuster
Drittprüfer:	Prof. Dr. Andreas Peichl
Tag der mündlichen Prüfung:	07. Juli 2021
Promotionsabschlussberatung:	14. Juli 2021

Acknowledgments

The process of writing a Ph.D. thesis can be a long journey, and it is not clear where the road will lead you. I was in the lucky position to have people who helped me to navigate through this adventure. Therefore, I would like to seize this opportunity to express my deep gratitude for their guidance and support.

First and foremost, I would like to thank my supervisor, Helmut Rainer. I am very grateful for his enthusiastic encouragement and continuous support throughout my entire time at the ifo Institute. Most of all, he gave me the freedom to discover and establish my own research interests and his thoughtful comments helped me to improve my research. I would further like to thank Stefan Bauernschuster for agreeing to be my second supervisor and Andreas Peichl for joining the examination committee. Furthermore, I would like to extend my warm gratitude to Natalia Danzer, who gave me valuable comments and feedback on my first chapter. In the process of getting this paper published in the Journal of Health Economics, I also benefited from constructive comments and helpful suggestions from the editor, Ana Balsa, and two anonymous referees. I am very thankful for the comments as they have helped improve the paper's quality considerably.

One chapter in this thesis is co-authored work. During this project, I have learned a great deal from my co-authors Maria Waldinger and Helmut Rainer. I am very thankful for the opportunity to work with them. Dominik Ammon, Maximilian Grasser, Paulina Hofmann, and Lina Leutner provided valuable research assistance over the years.

Next, I would like to thank the ifo Institute for the opportunity to enter a Ph.D. program in Economics, to study off-curriculum topics in instructive summer schools, and to present my research at many prestigious conferences both in Europe and the United States. Throughout most of my Ph.D., I also benefited from generous funding provided by the Leibniz Society, which is gratefully acknowledged.

I was fortunate to work with exciting data sets for all three chapters of my dissertation, and many people assisted me in this respect. When fighting the bureaucratic burdens to analyze the hospital registry data, I had strong support from the staff at the Research Data Centers of the Federal Statistical Office and the statistical offices of the Länder. I would particularly like to thank Heiko Bergmann and Nina Storfinger for their assistance in accessing the data. Furthermore, I would like to thank Helmut Becker from the Federal Criminal Office for his help in obtaining the crime register data.

My colleagues and other Ph.D. candidates at the ifo Institute and the MGSE have been very supportive, helpful, and fun over the last few years. The journey would have been less enjoyable without them. In particular, I would like to thank Daniel, Dennis, Eleonora, Fabian, Felix, Michael, Patrick, Stefan, and Victoria for helpful discussions and unforgettable experiences.

Lastly, I want to thank my family and good friends who have been an anchor during rough times and made good times even better. Most of all, I thank my wife Theresa and my children for their invaluable support, encouragement, and love. I am very grateful for having you and would not have achieved this without you.

Marc Fabel
Munich, March 2021

Contents

List of Figures

List of Tables

Preface

Almost 40 years ago, Edward E. Leamer (1983) criticized the state of applied econometrics at that time: *"Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analysis seriously."* (p.37). He urged applied economists to *"take the con out of econometrics"* and suggested to conduct more sensitivity analyses, which would test the robustness of the results when deviating from the preferred specification or functional form - an element that can be found in every applied paper today.

Nowadays, however, it seems that Leamer's critique is no longer justified as applied econometrics has experienced a *credibility revolution* over the last three decades (Angrist and Pischke, 2010). More emphasis on the quality of research designs has been the driving factor for the credibility revolution and the improved standards in the discipline as a whole. Identifying and estimating causal and policy-relevant parameters has been the bread-and-butter business of applied economists ever since. For more than 30 years, the *potential outcome* approach, sometimes referred to as the Rubin Causal Model, has been the standard framework to analyze causal problems in economics (Holland, 1986, Imbens, 2020).[1] For a while, experiments with random assignment (randomized controlled trials) seemed to be the only viable option for causal inference. However, randomized experiments are problematic or entirely not feasible in many instances, may it be due to financial, political, or ethical considerations (Athey and Imbens, 2017). For this reason, applied economists developed quasi-experimental designs since the late 1980s and have exploited natural experiments that allow to draw causal inference from observational data. Instead of relying on randomization through experiments, researchers need to find exogenous variation in the treatment variable, which is often induced by institutional rules (Van der Klaauw, 2014). Typical identification strategies to estimate causal effects in natural experiments are instrumental variables, differences-in-differences, and regression-discontinuity designs, amongst others. The change in empirical research was accompanied by an increase in the requirements on data. It is not only crucial to know the source of the exogenous variation but also to have a large enough sample of individuals affected by the natural experiment - requirements that are often not met by traditional surveys.

The credibility revolution would not have been possible with another contemporaneous transformation, namely a *data revolution*. Economic research has generally become much more empirical over the last decades. The share of published papers that are purely theoretical has been in decline, whereas papers that exploit self-collected data have been increasing (Hamermesh, 2013, Angrist *et al.*, 2017). Researchers have opened up new

[1] An alternative central framework for causal inference uses the work on *directed acyclic graphs*, which has received much attention in other disciplines, such as computer science, epidemiology, and other social sciences but not so much in economics (Imbens, 2020).

data sources, such as text data, private sector data, and administrative data (Gentzkow *et al.*, 2019, Langedijk *et al.*, 2019). Particularly important for estimating causal effects in the setting of natural experiments has been the improved access to large-scale administrative data, which offers substantial benefits (Einav and Levin, 2014). The (almost) universal population coverage allows that quasi-experimental designs can be applied in the first place. This is owed to the fact that quasi-experimental designs often investigate effects for populations at the margin of the natural experiment (e.g. birth cutoff rules). Furthermore, with detailed and accurate information, researchers can construct consistent indices to track outcomes over long periods of time. Lastly, in comparison to traditional surveys, administrative data does not suffer from selection problems or issues with missing data, such as attrition and non-response. Recently, there have been numerous advancements in the econometric literature to solve policy problems in large data sets with the help of machine learning methods (Mullainathan and Spiess, 2017, Athey, 2018, Huber, 2019). Promising applications in this respect are: estimating treatment effect heterogeneity, flexible modeling of treatment effects, and selecting from a high dimensional set of covariates.

This thesis exploits modern econometric techniques and unique large-scale data sets to identify and estimate causal and policy-relevant parameters. The first chapter investigates the causal impact of an expansion in post-natal maternity leave coverage on children's health outcomes in the long-run. The second chapter analyses the causal effects of professional football games on violent behavior. The third chapter examines the relationship between local strike participation in climate strikes and electoral outcomes.[2] The three chapters are self-contained and can be read independently. A consolidated bibliography can be found at the end of this thesis. The remainder of this section contains a brief summary of each chapter and a discussion on common features shared by the individual projects.

Chapter 1, *Maternity Leave and Children's Health Outcomes in the Long-Term*[3], assesses the impact of the length of maternity leave on children's health outcomes in the long-run. Previous literature examining the effects of leave schemes has focused on either short-run health or long-run educational attainment and labor market outcomes of children. This chapter aims to fill the gap in existing literature by investigating the causal effect of the length of maternity leave on children's long-run health outcomes. My quasi-experimental design evaluates an expansion in post-natal maternity leave coverage from two to six months, which occurred in the Federal Republic of Germany in 1979. The expansion came into effect after a sharp cutoff date and significantly increased the time working mothers stayed at home with their newborns during the first six months after childbirth. To establish causality, I leverage this cutoff date as a source of exogenous variation and employ a difference-in-differences approach to account for seasonal birth effects. In the analysis, I exploit German hospital registry data, which contains detailed information on the universe of inpatients' diagnoses from 1995-2014. By tracking the health of treated and control children from age 16 up to age 35, this study provides new insights into the trajectory of health differentials over the life-cycle. I find that the legislative change generated positive long-term health effects: My intention-to-treat estimates show that children born after the implementation of the reform experience fewer hospital admissions and are less likely to be diagnosed with mental and behavioral disorders - the most frequent diagnosis

[2]In contrast to the first two chapters, chapter 3 does not claim causality. The identification of causal effects is still work in progress at the time of submission.

[3]A version of this chapter has been published in the *Journal of Health Economics* (Fabel, 2021).

type for individuals in the observed age group. The results are driven by fewer hospitalizations among men and are more pronounced for individuals in their late 20s and after. Furthermore, exploiting the fine granularity of the diagnoses codes, the results show that treated males experience fewer mental and behavioral disorders due to the use of psychoactive substances and fewer incidences of schizophrenia. The results imply substantial annual health cost savings of about 6.6 million euros per birth cohort, just for the reduction in mental and behavioral diseases alone. The chapter suggests that economically significant benefits of maternity leave may be realized with quite a lag and in different areas than what policymakers had in mind initially. Consequently, cost-benefit analyses could come to wrong conclusions if the vast effects on children's long-run health outcomes are not accounted for.

Chapter 2, *Crime Externalities from Football Games in Germany*, investigates the causal impact of professional football games on physical violence in Germany. I employ a generalized difference-in-differences strategy that exploits the variation in the timing of the football games. I compare the number of assaults on days with and without football matches while controlling for date heterogeneity, weather, and holidays. To examine the effect of football games on assaults, I match geo-coded information on almost 4,500 games with comprehensive crime registry data for the period 2011-2015. My results suggest that a home game leads to a 21.5 percent increase in the assault rate. The effects are primarily driven by male victims who do not know the perpetrator and decrease with the victims' age. When investigating channels, I do not find evidence that emotional cues from unexpected or unsettling events during the game would unleash physical violence. In fact, there is suggestive evidence for other potential channels: spectators mimic the behavior displayed on the field, violent viewers self-select into specific matches, and a pure agglomeration effect coming from spectators. My results show that the external effects of football games on violent behavior are large in magnitude and economically relevant. To be precise, professional football games explain almost 18 percent of all assault reports in the affected regions and generate annual social costs of 95 million euros. By quantifying the criminal externalities in the setting of professional football in Germany, this paper contributes to the widely noticed debate on the reimbursement of costs associated with high-risk professional football games.

Chapter 3, *The Power of Youth: School Strikes for Climate and Electoral Outcomes*, which is joint work with Helmut Rainer and Maria Waldinger, analyses the relationship between local participation rates in climate strikes, also known as Fridays for Future, and electoral results from four elections in Germany in 2019. In that year, large crowds of protesters, primarily students, gathered for weekly climate strikes to demand far-reaching, rapid, and effective measures to mitigate climate change. Alone during the Global Week of Climate Action in September 2019, more than 1.4 million people demonstrated in 550 places across Germany. In the first part of the empirical analysis, we propose a local measure of strike participation using cell phone-based tracking data. With the help of a rich fixed-effects model, we remove the predictable variation in the number of journeys between origin-destination pairs. The residualized journeys are matched with geo-coded locations of more than 2,300 strikes and indicate how many participants attend a strike and where they come from. In the second part of the empirical analysis, we investigate the relationship between climate strike participation levels and electoral outcomes. Using a first-differences specification, we show that municipalities with higher strike participation rates are more likely to vote for the Green Party. Additionally, voter turnout is increasing with increasing levels of strike participation. Our results are robust to alternative measurements of strike participation. In economics, an extensive literature studies the intergenerational transmission of values and political beliefs from parents to children. The results of this study suggest that there may be a reverse transmission mechanism, namely that children have an impact on their parent's environmental awareness and preference for green politics.

The three chapters are the result of distinct research projects that I carried out during my time as a Ph.D. student and, as such, differ along many dimensions. However, there are common characteristics that are shared by all chapters throughout the dissertation, such as the exploitation of novel large-scale data sets and clean identification strategies. To be precise, the first chapter uses 20 years of hospital administrative data covering the universe of inpatient cases in Germany. The empirical approach compares health outcomes of children born around a cutoff date in the reform year to the outcomes of unaffected children born around the same threshold but one year before the reform. The control group serves the purpose of netting out seasonal effects and represents the counterfactual scenario, i.e. what difference in health outcome would be expected around the threshold in absence of the reform. This chapter benefits particularly from the aforementioned advantages associated with administrative data. The local identification aspect requires a large number of observations at the margin of the natural experiment (i.e. individuals born around the cutoff date), which the hospital administrative data can easily provide with 20 million observations per cross-section. Furthermore, the consistent coding of diseases according to the World Health Organization's classification system (ICD) allowed the tracking of health differentials over two decades. The second chapter draws on crime register data from the Federal Criminal Police Office, which contains more than 4.8 million individuals that became victims of crimes against their legally protected personal rights over the five-year sample period. I enrich the register data containing precise information on the time and the location of offenses with self-collected data from the web. I went to great lengths to scrape information on detailed game statistics and pre-game betting odds for almost 4,500 matches. After merging the data sources, I identify the causal impact of a football game by employing a generalized difference-in-differences strategy. The counterfactual rate of physical violence is obtained using the local assault rate in weeks where no game is scheduled. The third chapter utilizes cell phone-based tracking data to measure local climate strike participation rates. The data is obtained from the private sector and translates mobile signals from customers of one of the largest carriers in Germany into a daily number of journeys between two geographies. In a rich fixed effects model, we use 64 billion trips to remove the predictable variation in the number of journeys between origins and destinations.

In the spirit of the credibility and the data revolution, this dissertation presents results using exciting novel data sets and transparent empirical strategies to address policy-relevant research questions. I hope this dissertation will spark more debate and research on the topics covered in the three chapters.

Chapter 1

Maternity Leave and Children's Health Outcomes in the Long-Term*

1.1 Introduction

Over the last decades, family leave programs have been strongly promoted in large parts of the Western hemisphere. The average length of paid maternity and parental leave across all OECD countries rose from 14.0 weeks in 1970 to 55.4 weeks in 2016 (OECD, 2020). Leave policies allow parents to take a break from work and focus on child care. These leave schemes are based on evidence showing that the first year in a child's life is essential for subsequent child development (Currie and Almond, 2011). The impact of family leave schemes goes beyond protecting against job dismissal and compensating income losses. They have gained significant momentum as a policy tool in encouraging female employment (Blau and Kahn, 2013) and fertility (Lalive and Zweimüller, 2009), advocating for gender equality (Kotsadam and Finseraas, 2011), and safeguarding the well-being of mother and child (Bütikofer *et al.*, forthcoming). To achieve these goals, the German Federal Ministry of Family Affairs is expected to pay 7.25 billion euros for parental leave in 2020.[1] Previous literature examining the effects of leave schemes has focused on either short-run health or long-run educational attainment and labor market outcomes. However, evidence on the effect of family leave policies on children's health outcomes in the long-run remains scarce.

In this paper, I aim to fill the gap in existing literature by assessing the impact of the length of maternity leave (ML) on children's long-run health outcomes. My quasi-experimental design evaluates an unexpected expansion in post-natal ML coverage from two to six months, which occurred in the Federal Republic of Germany in 1979. The expansion came into effect after a sharp cutoff date and significantly increased the time working mothers stayed at home with their children during the first six months after childbirth. To estimate the causal effect of the length of ML on child health outcomes, I exploit the exogenous variation stemming from the reform, which provides a treatment assignment that is plausibly random. All previously employed women who gave birth on or after May 1, 1979 were eligible for six months of ML after childbirth. On the contrary, mothers who delivered their baby

*A version of this chapter has been published in the *Journal of Health Economics*, Vol. 76, 102431, in March 2021, doi: `https://doi.org/10.1016/j.jhealeco.2021.102431`.

[1]This amount corresponds to just above 2% of the entire 2020 federal government budget (Federal Ministry of Finance, 2020).

before the cutoff date were entitled to only two months of job-protected ML. In order to account for potential seasonal birth effects, I employ the following difference-in-differences approach: I compare health differentials between children born in the months before and after the implementation of the reform, with health differentials between children born in the same calendar months but in the previous year, when no legislative change took place.

I exploit German hospital registry data containing detailed information about the universe of inpatient cases for the years 1995 to 2014. By tracking health outcomes of individuals aged 16-35 who, as children, were (or were not) exposed to the reform, I find significant and robust evidence that the ML expansion improves children's health over the life-cycle. Children subject to the more generous leave scheme are on average 1.7 percent less likely to be hospitalized. The results are driven by fewer hospital admissions among men and are stronger for individuals in their late 20s and after. Additionally, using the diagnoses codes of hospitalizations, the results show that the largest driver for the decline in hospital admissions comes from a reduction in mental and behavioral disorders (MBDs). The effects on MBDs mirror the overall results well, particularly with respect to age and gender differences. MBDs not only bear the largest relative contribution in the reduction of hospitalizations but they also resemble the most frequent diagnosis type for individuals in the age group that are observed.[2] Because of that importance in terms of effect size and prevalence, I exploit the fine granularity of the diagnoses codes and study the effect of the 1979 ML reform on subcategories of MBDs. I find that treated males experience fewer MBDs due to the use of psychoactive substances and fewer incidences of schizophrenia. These results are consistent with Canetti *et al.* (1997) and Enns *et al.* (2002) who show that parental bonding, in particular with the mother, is linked to the development of mental disorders later in life.[3] The results are robust to alternative specifications of the dependent variable and the level of aggregation, different estimation approaches, and also various placebo tests.

In order to elicit the possible underlying channels of the relationship between the ML expansion and children's long-run health outcomes, I compare the impact of the reform on children's health outcomes in rural and urban areas, as mothers in urban areas had a higher propensity to work in 1979. The results show that the overall effects on hospitalization and MBDs are driven mainly by urban areas. A greater labor force participation implies that more mothers were eligible for ML and as a consequence, more children were affected by the reduction of maternal labor supply. Taken together, the results suggest that the reform had a larger impact on children's health outcomes in regions where there was a stronger reaction to the 1979 ML reform.

The notion that later-life outcomes originate from early childhood is not novel. Barker (1990) postulated that conditions in-utero and during infancy have long-lasting effects on later life health. Early experiences may influence adult physical and mental health in two ways (Shonkoff *et al.*, 2009). On the one hand, there is a cumulative process at play, in which early experiences trigger a repetitive provocation of neurobiological responses that may become pathogenic. On the other hand, the environment during key developmental stages is biologically embedded into regulatory physiological systems such that it can impact adult disease and risk factors latently. In these sensitive periods, the developing

[2]Over the entire time frame, MBDs account for one-third of the hospital reductions. For individuals aged 32-35, almost 50% of the reduction in hospitalizations is stemming from the decline in MBDs.

[3]Enns *et al.* (2002) show that *"experiences with one's mother were more consistently associated with adult mental disorders [...] However, there appeared to be some unique effects for externalizing disorders (substance use disorders and antisocial personality disorder) in males"*.

brain's architecture is modified considerably and is particularly sensitive to environmental stimuli. In both cases, the effects of experiences made in early life may be latent initially until the onset of a particular condition.

My results show that the 1979 ML expansion entails benefits that were not at the heart of policy debates that surrounded its implementation. Yet, back-of-the-envelope calculations based on my estimates indicate that the health effects of the ML expansion are substantial. For instance, the reform leads, on average, to 370 fewer diagnoses of MBDs for one birth cohort per year. This implies health cost savings of about 6.6 million euros per birth cohort per annum, based on an estimated cost of 17,850 euros per MBD diagnosis.[4] The effects on MBDs are particularly interesting as this disease category is the most prevalent among the age group under consideration (15-35 years). In addition, MBD inpatients have on average the longest length of stay in comparison with other disease types.[5] For these reasons, MBDs are among the most expensive disease categories.

This paper complements various strands of the literature. At a very general level, it relates to studies that try to explain the role of early childhood experiences in later life outcomes.[6] The majority of studies investigating the long-run impacts of ML, however, focus on human capital accumulation and labor market outcomes and typically have two common findings: positive effects of leave duration on child outcomes, or no effects at all.[7] For instance, it has been documented that extending leave leads to higher cognitive skills (Albagli and Rau, 2018), better PISA test scores (Danzer and Lavy, 2017), lower dropout rates from high school, and higher earnings of individuals when aged 30 (Carneiro *et al.*, 2015). On the other hand, some studies have found no effects on children's test scores and the propensity to graduate from high school (Dahl *et al.*, 2016), or on children's years of schooling, wages, and the likelihood of full-time employment (Dustmann and Schönberg, 2012). Second, this article relates to the literature on the impact of ML on short-run health outcomes. In contrast to the previous literature field, the results in this strand of the literature are less ambiguous. Expanding leave mandates have been shown to have a positive impact on children's health in the short-run.[8] ML extensions improve infants' overall health and reduce asthma rates (Bullinger, 2019), decrease the incidence of emotional disorders (Sayour, 2019), reduce the likelihood of early-term birth, and increase newborns' birth weights (Stearns, 2015).

This paper aims at combining the above-mentioned two strands of literature by providing causal estimates of the length of ML on child health outcomes in the long-run. Two studies relate closely to this paper. First, Danzer *et al.* (forthcoming) evaluate an Austrian reform from 1990 that increased paid and job-protected parental leave from 12 to 24 months. They find positive effects on children's disability status up to age 23 and fitness

[4]The estimated social cost of 17,850 euros per MBD diagnosis is obtained by using data on the cost of illnesses by disease for the age group 15-45 in 2015, as provided by Federal Statistical Office (2015).

[5]The average length of stay for MBD diagnoses is 20.1 days in comparison to a general average of 7.6 days (Federal Statistical Office, 2012, p. 5).

[6]Currie and Almond (2011) and Almond *et al.* (2018) offer detailed summaries of recent work in this strand of research.

[7]Other studies investigate different features of leave schemes such as the benefit levels. For instance, Ginja *et al.* (2020) find in the Swedish context that higher levels of parental leave benefits improve children's educational outcomes.

[8]The results of Beuchert *et al.* (2016) are a notable exception. They exploit a reform of the parental leave scheme in Denmark and find null effects on children's health outcomes in the first three years after birth. The Danish reform increased post-birth ML by, on average, 32 days. The length of paid leave was increased from 24 weeks (14 weeks maternity and 10 weeks of joint leave) to 46 weeks (14 ML + 32 joint).

for military service (males only). Yet, they do not observe any effects of the Austrian reform on children's labor market outcomes. The authors also show effect heterogeneity by counterfactual mode of care. Their analysis demonstrates that the positive effects are driven by regions with no nurseries, i.e. regions where informal care is substituted with maternal care. Second, Ahammer *et al.* (2020) examine an Austrian reform from 1974, which expanded prenatal ML from six to eight weeks. They find no effect on outpatient expenses and hospital days for individuals in Upper Austria when aged 25-40.

The empirical analysis in this paper expands previous literature on several dimensions. In contrast to previous literature, this study is able to investigate the effect of a ML expansion on an unusually broad range of health outcomes. Using the universe of inpatient cases that were discharged from a hospital or rehabilitation facility, I examine the effects of the 1979 ML reform on the patients' main diagnosis, which is coded according to the WHO's classification catalog. This way, I can investigate the effect of a ML expansion on many relevant disease categories. Moreover, the institutional setting of Germany and the data allow me to trace out the trajectory of health differentials across 20 years of children's adulthood (from age 16 up to age 35). The longitudinal structure of the data permits a long-run perspective as children and hence, differentials are allowed to develop over time.

The remainder of the paper is structured as follows. The next section provides information about the 1979 ML reform. Section 1.3 explains the data and variables. Section 1.4 discusses the empirical design. Section 1.5 reports results. Section 1.6 contains a discussion of the conceptual framework and mechanisms. Section 1.7 concludes.

1.2 Background

1.2.1 Institutional Set-Up

In contrast to the United States, ML laws have been in existence much longer in Germany.[9] Since the mid-1950s, employed mothers held the right to a paid protection period of six weeks before and eight weeks after childbirth, during which they were not allowed to work.[10] During this so-called 'mother protection period', women were protected from being dismissed and upon their return to work they had the right to be placed to a job comparable to their prior assignment. The benefits in this period corresponded to a 100% replacement rate and was equivalent to women's average income over the three months before childbirth.[11] This pre-reform setting is to some extent comparable to the current maximum of 12 weeks of unpaid, job-protected leave in the US (under the FMLA) and the minimum of 14 weeks of paid, job-protected leave in the EU.[12]

[9]The following facts about ML and benefit legislation are based on information by Ondrich *et al.* (2002), Schönberg and Ludsteck (2014), Dustmann and Schönberg (2012), and Zmarzlik *et al.* (1999). The leave scheme described here does not correspond to the current system, which has been in place since 2007. Kluve and Tamm (2013) offer a good overview of the present leave regulations.

[10]Before another reform took place in 1986, only mothers were eligible for job-protected leave.

[11]The payment was co-funded by public health insurance funds (750 Deutschmarks (DM) per month), the federal government (400 DM, one-time payment) and employers (the remainder).

[12]Since the Family and Medical Leave Act of 1993 (FMLA), mothers in the US have been entitled to leave if they worked for at least one year with their employer, accumulated a minimum of 1250 working hours during that year, and if they worked for an employer with at least 50 employees (Baum, 2003).

In 1979, the socio-liberal coalition of chancellor Helmut Schmidt passed a reform bill, which introduced four extra months after the end of the mother protection period. In other words, the total length of ML after childbirth (job protection and benefits) increased from eight weeks to six months (see Appendix Figure A.4).[13] The federal government wanted primarily to safeguard maternal health after childbirth with the reform. However, positive spill-over effects on the child were acknowledged.[14] While the initial benefits of the period from six weeks before and eight weeks after childbirth did not change, the payments were equal to 750 DM from the third month after delivery. This amount corresponded to approximately 44% of average pre-birth earnings in 1979 (Schönberg and Ludsteck, 2014). Although eligibility for ML was universal among working women, the approximated take-up rate was low where only about 45% of mothers took advantage of the ML reform in 1979 (Dustmann and Schönberg, 2012).

The reform was initiated by a draft bill on January 5, 1979. The final law was ratified by the German Bundesrat (the Upper House of the German Parliament) on May 19 and by the German Bundestag (the Lower House) on June 22, 1979. All previously employed women who gave birth on or after May 1, 1979 were eligible for six months of ML after childbirth. In contrast, mothers who delivered their baby before the cutoff date were entitled to the 'common' two months of job-protected ML. It is noteworthy that strategic conceptions were impossible due to the short period between the draft bill and when the reform took effect. This implies that families in the sample of analysis could not adjust their fertility behavior in response to the reform and the 1979 ML extension can and should therefore be seen as a quasi-experiment. This issue is discussed in great detail in the Appendix.

1.2.2 Female Labor Force Participation and Childcare Situation

In April 1979, around 38% of the German labor force was comprised of women and almost every second woman aged 15 to 65 years (49.7%) was active in the labor market (Federal Statistical Office, 1981).[15] Yet, there was pronounced heterogeneity in female labor force participation rates. For instance, female labor market attachment was at 62.4% for singles, 45.2% for married, 32.5% for widowed, and 76.5% for divorced women. Additionally, there was a strong gradient with respect to age. While 69.2% of all women aged between 20 and 25 years participated in the labor force, the share was 55.0% for women in the 30-35-year-old age bracket.[16] The high numbers for younger and single women indicate that a high share of mothers-to-be was an active part of the labor force and thus eligible for ML.

Nevertheless, in addition to the female labor force participation rate, the counterfactual mode of care had an impact on how the reform can alter children's outcomes. Danzer *et al.* (forthcoming) show that the 1990 parental leave expansion in Austria had a positive effect on children in regions where there was no formal childcare. In other words, there were only positive effects of the reform in cases when informal child care was substituted with parental care. Hank and Kreyenfeld (2001) describe the childcare situation in West Germany in the

[13] Refer to: *'Gesetz zur Einführung eines Mutterschutzurlaubes'* (Maternity leave law), Bundesgesetzblatt (Federal law gazette), Part I, Nr. 32, p.797-802, 30.06.1979.

[14] Refer to: *'Gesetzesentwurf der Bundesregierung'* (Draft bill), Drucksache 8/2613.

[15] As a comparison, 50.9% of all women (16+) in the US participated in the labor market at the same time (see US Bureau of Labor Statistics).

[16] 84% of all children were delivered by women aged 20 to 35 (Federal Statistical Office, 1981). In this context, it is therefore relevant for the study to focus on this age group.

late 1970s as a *'patchwork [of] childcare arrangements'*, meaning that parents had to rely on a broad range of care types, such as parental care, daycare centers, social networks, and private childminders. The different forms of childcare, however, varied greatly in cost and quality. In 1980, only 1.5% of children attended a public *'Krippe'* (nursery for children aged 0-3) (Konsortium Bildungsberichterstattung, 2006, p. 34).[17] Since public daycare was virtually non-existent, parents had to rely almost exclusively on informal care, apart from parental care. Thus, the situation in the Federal Republic of Germany in the late 1970s allowed for a substitution from informal arrangements to maternal care.

1.2.3 What We Already Know about the 1979 Maternity Leave Reform

To date, two studies by Dustmann and Schönberg (2012) and Schönberg and Ludsteck (2014) have examined the effects of the 1979 ML reform.[18] Overall, both papers show that the reform had a large impact on mothers' labor market outcomes, particularly in the short-run.

First, many mothers adjusted their labor supply downwards during the four months of extra leave and returned to the labor market as soon as the leave period terminated. The long-run maternal labor supply (the time period beyond six months after childbirth), however, was less affected. For instance, while the reform decreased the share of mothers who returned to the labor market by the third month after childbirth by 30.5 percentage points, the corresponding reduction at the 52nd month after childbirth is around one percentage point. In total, the change of postnatal ML from two to six months caused mothers to postpone their return to work by, on average, 0.835 months.[19] Approximately two-thirds of the decline in short-run female labor force participation was the result of a contraction in full-time work.

Second, the ML expansion led to changes in mothers' income. The overall increase in average cumulative total income was 1,700 DM.[20] There were two effects on average available income. On the one hand, there was a decline in available income as mothers returned to work later because of the reform and received 750 DM, which corresponded to only 55% of mothers' average post-birth wage. On the other hand, there was an increase in available income for mothers who would have stayed at home even without the reform. The second effect, which is a crowding out of unpaid leave, dominates the first one, such that there is an overall increase in available income. The impact of the reform on cumulative income varied substantially, depending on the position in the wage distribution: Mothers in the lowest tercile of the wage distribution had 2,850 DM of additional income, while the increase for women in the highest tercile amounted to 1,050 DM.

[17] In the 1970s, the provision of part-time care for pre-schoolers (4-6 years) was established. However, parents did not have the legal entitlement to a slot in a public *'Kindergarten'* until 1996. For toddlers (1-3 years), parents had a legal claim to a care slot in a *'Krippe'* since 2013.

[18] Both, Dustmann and Schönberg (2012) and Schönberg and Ludsteck (2014) analyze the impact of the ML expansion on maternal labor market outcomes. While the former study evaluates additionally changes in child outcomes due to the reform, the latter focuses on maternal labor market outcomes and elicits more in-depth results.

[19] This number corresponds to the number of months away from work in the first 40 months since childbirth.

[20] Maternal cumulative total income is defined as the accumulated total income up to the point when the child is 40 months old. It consists of monthly earnings when the mother is working, equals to the benefits when she is on leave, and is zero otherwise.

Although there were distinct effects on maternal labor market outcomes, particularly in the short-run, Dustmann and Schönberg (2012) do not find evidence that the reform had an impact on children's educational attainment and labor market outcomes. There was no effect of the reform on years of education, wages, or the share of individuals in full-time employment.

1.3 Data

I use hospital register data spanning the period from 1995 to 2014, provided by the Research Data Centers of the Federal Statistical Office and the statistical offices of the Länder.[21] The register contains information on the universe of German inpatient cases; in the 2014 cross-section this amounts to 19.6 million observations. The administrative data covers *all* patients that were discharged from *any* hospital or medical prevention/rehabilitation facility in Germany in each reporting year.[22] Unless otherwise noted, I restrict the sample to individuals belonging to either treatment or control cohort, defined as individuals born between Nov 1978–Oct 1979 and Nov 1977–Oct 1978, respectively (see section 1.4 for details). The data includes the patient's main diagnosis, the length of stay, whether the patient died or underwent surgery, and in which medicating specialist department the patient stayed the longest. Furthermore, the register contains socio-demographic characteristics such as month and year of birth, gender, and postal code of the place of residence.

The main diagnosis indicates the major reason for the patient's hospital admission. It is coded according to the guidelines of the 'International Statistical Classification of Diseases and Related Health Problems' (ICD), which are maintained by the World Health Organization (WHO).[23] Up to and including 1999, the coding had followed the ICD-9 classification, since 2000 the ICD-10 system has been in place.[24] Table A.2 in the Appendix provides a summary of the frequencies of diagnosis categories. These categories are called chapters and together constitute the hospitalization variable.[25] In the pooled sample, "Injury, poisoning and certain other consequences of external causes" are the most frequent diagnosis types, followed closely by MBDs, and diseases of the digestive system. A similar pattern is observed in the cross-section: Appendix Figure A.5 gives an overview of the five most common diseases and health problems across age brackets in 2014. For the age group observed in the sample of analysis who are people between 15 and 35 years, MBDs are the most common diagnoses (357 thousand diagnoses), followed by injuries (310 thousand diagnoses), and diseases of the digestive system (260 thousand diagnoses).

Figure 1.1 illustrates trends in hospital admissions for the treatment and control cohort from age 17 through 35. Panel A shows an S-shaped line for the number of admissions. The rate of hospitalization increases until age 19 (35,000 cases per year), and then decreases

[21]Due to data confidentiality regulations, data access was provided on-site at the research data center.

[22]The data does not cover hospitals of the penal system and police hospitals. Military hospitals are included to the extent to which they offer services to civilians. Medical prevention/rehabilitation centers have been included since 2003 if they have more than 100 beds.

[23]Please note that the data exploits a German modification that is issued by the German Institute of Medical Documentation (DIMDI) - a subordinate authority of the Federal Ministry of Health.

[24]The numbers of the ICD classification refer to the revision, which is in place at the year of reporting. The ICD-9 classification is a 3 digit numeric code, whereas ICD-10 uses a 4 digit alphanumeric code.

[25]I exclude diagnoses related to "pregnancy, childbirth and the puerperium" and others that occur infrequently.

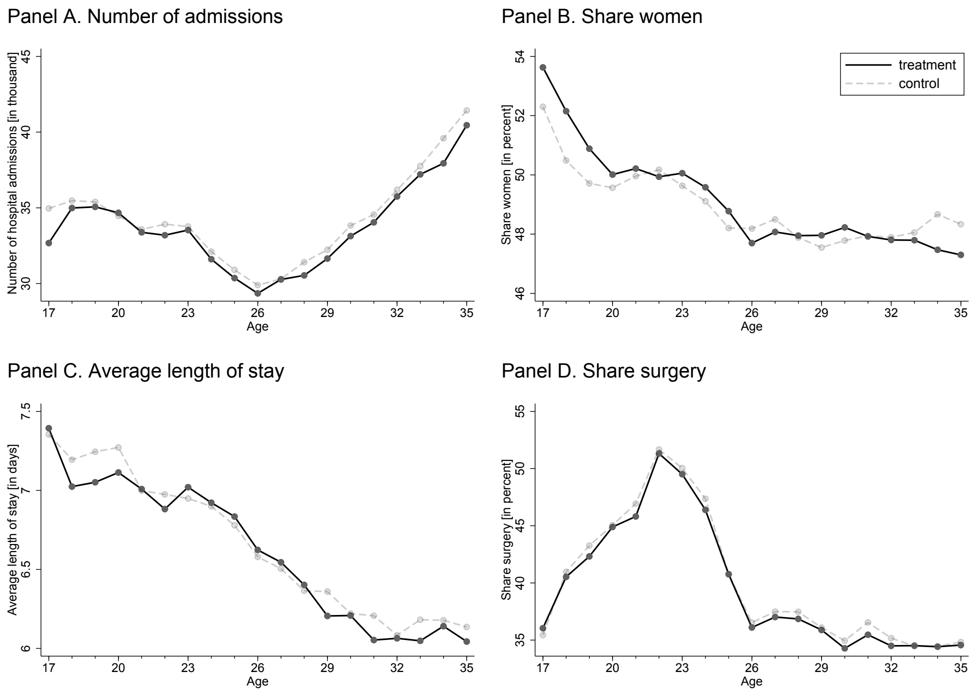


Figure 1.1. Hospital admissions

Notes: The figure depicts the evolution of key variables for the treatment and control cohort (only pre-threshold months, i.e. for individuals born between November 1977 and April 1978 as well as November 1978 and April 1979, respectively) over the ages from 17 to 35. The dark lines correspond to the treatment cohorts, whereas control units are marked by light dashed lines. Hospital admissions are defined as the sum of all diagnosis chapters listed in Panel A of Table A.2.

up to the age of 26 after which the number of admissions grows again (in 2014, there are 40,500 admissions). In panel B, it can be seen that the share of women decreases from around 55% with the age of 17 to below 48% when aged 35. Panel C shows a reduction in the average length of stay from 7.7 days to 6 days from the beginning to the end of the observation period. Panel D shows a hump-shaped evolution of surgeries that are related to hospitalizations. The share of surgeries increases up to the age of 22, and then declines until the age of 26. After that age, the share of inpatients with a surgery remains constant at around 35%.

For the analysis, I aggregate the number of diagnoses per chapter by birth month, birth year, and reporting year, and define the outcomes as the number of diagnoses per 1,000 individuals.[26] I use all hospitalizations of treatment and control individuals who reside in the area of the former Federal Republic of Germany.[27] The DiD baseline specification is therefor made up of a pseudo-panel with $2 \times 12 \times 20 = 480$ (cohorts$\times$months-of-birth$\times$reporting years) observations. In order to abate any confounding effects that might be triggered by

[26]In my baseline specification, I use the monthly number of births in the denominator (based on Federal vital statistics). In the robustness section, I present results with the approximated number of current inhabitants on different regional levels. The advantage of using the original number of births is twofold. First, it allows the tracing out of differentials over a longer period since the population data is only available from 2003 onwards. Second, I avoid inducing measurement error in the dependent variable as there is only information on the number of individuals aged x years. In order to obtain an approximate number of persons per birth-month, I multiply the number of people per birth-year with month-of-birth weights coming from either the German Micro Census or the original fertility distribution.

[27]Since Berlin cannot be assigned to either FRG or GDR unambiguously, it is dropped from the analysis.

differential maturity between treatment and control group, I compare the two birth cohorts at the same age. To achieve this, a control observation is shifted from period t to period $t+1$, which decreases the number of effective observations to 456. The shifting reduces the number of observations at the beginning and the end of the observed time frame. For instance, in 1995, the treatment (control) cohort is aged 16 (17). As there is no control group that is aged 16, I drop these 12 months of the treatment group from the sample. The control group that is already 17 years old is shifted to 1996. Analogously, I drop the control cohort in 2014.

The data set has three main advantages. First, compared to survey data, the hospital registry data covers the universe of German inpatient cases and is consequently not prone to sampling errors or problems associated with attrition. Additionally, the large number of patients provides sufficient statistical power to identify local effects. Second, due to the longitudinal character of the data set, this study is able to trace out the trajectory of children's health differentials over 20 years of their adulthood. Third, measurement error is unlikely to be present in the data on health outcomes. As the diagnosis code matters for remuneration, the ICD code is of high quality. Moreover, in comparison with self-reported survey data, administrative data does not suffer from issues related to social desirability bias.

Nevertheless, the data set has various drawbacks. First, the source of data limits the analysis to relatively severe health events, which are typically encountered in the context of hospitals and medical prevention/rehabilitation centers. Yet, some health conditions, which are diagnosed elsewhere, for instance at a general practitioner, could be more impacted by the reform than the diagnoses observed. Second, although the registry data is rich in both the number of cases and the quality of its entries, it contains only a few socio-economic variables. This implies that apart from the average effects of the 1979 ML reform, I am only able to perform heterogeneity analyses for selected subgroups. Lastly, there is no information about the place of birth, implying that the region in which a patient is observed does not necessarily coincide with the patient's place of birth. It would be ideal to exclude individuals from the analysis whose mothers were not affected by the reform, such as foreign-born children and children born in the German Democratic Republic (GDR). Yet, as I am unable to do so, the intention-to-treat (ITT) estimates are reduced in size due to the infeasibility of separating unaffected from affected individuals. The implicit assumption for unbiased estimates is that migration to the area of the former FRG occurred at random with respect to the month-of-birth. To alleviate this concern to some extent, I aggregate the baseline specification to the level of the former FRG and GDR and use more disaggregated data only for the investigation of effect heterogeneity and robustness tests. In doing so, I limit the impact of within region migration in the area of the former FRG or GDR.

1.4 Empirical Strategy

In order to estimate the causal effect of the length of ML, I exploit the 1979 reform's eligibility rule, which is contingent on children's birth date. Children born on/after the specified birth cutoff date May 1, 1979 fell under the new regime, during which their mothers were eligible for six months of ML after childbirth. Mothers of children born before the threshold were entitled to two months of postnatal leave. Assignment to one of the two schemes is a deterministic function of the birth date of the child.

A regression discontinuity design (RDD) might constitute a first potential identification strategy, in which one compares health outcomes of children born around the threshold. The identifying assumption is that the children on both sides of the cutoff are on average comparable, with the only notable exception that their mothers were entitled to different lengths of ML. Table A.3 contains estimates from an RDD with a linear polynomial and different slopes on both sides of the cutoff. Reassuringly, the direction and the magnitude of the RDD estimates match the corresponding estimates from my preferred estimation approach closely. However, the large standard errors illustrate the resulting precision problem associated with discontinuity designs in this context. The lack of precision is rooted in the fact that the birth date is only available at the monthly level.

The aggregation at the birth month level also entails other challenges. A large body of literature suggests a strong relationship between season of birth (SOB), health, and other socioeconomic outcomes.[28] The estimated health differentials of children born before and after the reform date could be biased if these SOB effects are not accounted for. The estimated effect may be partly driven by the difference in health outcomes stemming from the seasonality component rather than by the ML expansion itself. A difference-in-discontinuities design could account for the SOB effects but suffers from the above-mentioned precision problems. For this reason, I use the following difference-in-differences (DiD) approach. Children born around May 1, 1979, the year of the reform, constitute the treatment cohort, while children born around May 1, 1978 are part of the control cohort.[29] I compare differences in health outcomes of treated children born before and after the reform cutoff date to differences in health outcomes of control children born around the same threshold, one year prior to the reform. This accounts for the seasonality component while preserving the local identification aspect. The implicit identifying assumption is that seasonality is time-invariant. In other words, treatment and control group share the same SOB effects.

The main specification to estimate the effect of the length of ML on children's health outcomes corresponds to the following equation:[30]

$$\mathrm{Y}_{mt} = \gamma_0 + \gamma_1 \mathrm{Treat}_m + \gamma_2 \mathrm{After}_m + \gamma_3 (\mathrm{Treat}_m \times \mathrm{After}_m) + \psi_m + \rho_t + \varepsilon_{mt} \tag{1.1}$$

[28] The seasonality may come about due to reasons that are associated with either pre- or postnatal factors. First, the seasonality might arise due to selective conception, i.e. the socioeconomic composition of mothers varies over time (Buckles and Hungerman, 2013). Second, Currie and Schwandt (2013) argue that SOB effects may come from seasonal patterns of in-utero disease prevalence (e.g. influenza) and nutrition. Lastly, the seasonality with respect to time of birth may also be the result of postnatal social factors such as age-based cutoff rules at school-entry (Black *et al.*, 2011).

[29] Dustmann and Schönberg (2012) use in total three birth cohorts as control groups, two cohorts before and one cohort after the treatment cohort: group 1 born 11/1976-10/1976, group 2 born 11/1977-10/1978, and group 3 born 11/1979-10/1980. I choose individuals born one year before the reform as control group for the main specification. This is done for two reasons. First, with more cohorts as control groups, it may be less likely that the identifying assumption (time-invariance of seasonality) holds. Second, taking a birth cohort in the year after the policy change as control group might invalidate the comparability between treatment and control group, as parents might have had enough time to adjust their fertility behavior. Nevertheless, I present results with the addition of more control cohorts in the robustness section.

[30] The estimation procedure can also be found in similar contexts in Lalive and Zweimüller (2009), Dustmann and Schönberg (2012), Ekberg *et al.* (2013), Schönberg and Ludsteck (2014), Lalive *et al.* (2014), Danzer and Lavy (2017), Avdic and Karimi (2018), and Huebener *et al.* (2019).

where Y_{mt} is the number of diagnoses per thousand individuals of the cohort born in month m, at time t. The treatment cohort is represented by Treat_m, a dichotomous variable equal to one for groups that are born in the months before and after the legislation change, and zero otherwise. The analysis presents results for different estimation windows around the threshold date. In the widest specification, the treatment cohort includes children born between November 1978 and October 1979, implying a bandwidth of half a year around the cutoff. After_m is a dummy variable that is equal to one if individuals are born in the month of May and after, i.e. born in May-October in the widest specification, for both treatment and control cohorts. ψ_m, ρ_t are month-of-birth and reporting year fixed effects, respectively. Initially, Y_{mt} corresponds to outcomes observed over the pooled time period between 1995 and 2014. Subsequently, I break up the entire time frame in different age groups and apply a life-course approach by running the regression for each year of life separately. The parameter of interest is γ_3, which captures the effect of the policy change on health outcomes. As there is no information on whether the children's mothers were on leave, the identified parameter should be taken as the intention-to-treat effect.[31]

Standard errors are clustered at the birth month $\times$ birth year level to account for the likely correlation of the error ε_{mt} over time for a given month of birth cohort.

The Appendix covers potential threats to the validity of the study design, both at the cutoff date (self-selection into treatment) and across the distribution of birth months (seasonality and age of school entry effects). An examination of the number of births around the threshold reveals no evidence that parents strategically delayed births in the reform year, which lends support to the use of the 1979 ML reform as a valid natural experiment. Nonetheless, to rule out the possibility that strategic changes in fertility and delivery pose a threat to the identification strategy, I test the robustness of the results by applying a *'Donut'* specification, in which I exclude children who were born in the month before and after the policy was implemented.

1.5 Results

1.5.1 Hospital Admissions

Table 1.1 reports DiD estimates of the impact of the ML expansion on hospital admissions, using the specification in equation 1.1. The dependent variable is the hospitalization rate, which is defined as the annual number of hospital admissions per 1,000 individuals.[32] Panel A presents ITT results based on the pooled data for the years 1995 to 2014 (age 17 to 35). The baseline coefficient in column 1 suggests that the ML reform significantly reduces the fraction of annual inpatient treatments in hospitals by an average of 2.1 cases per 1,000

[31] The ITT effect identifies the causal effect of being *assigned* to treatment, which is labeled as the reduced form in an instrumental variables setting. To get to the local average treatment effect, the effect on compliers, the ITT is divided by the first stage (Angrist and Pischke, 2009). To give the resulting Wald estimand a causal interpretation, you need to assume, next to a standard monotonicity assumption, that the 1979 ML reform affected all determinants of child development only through the reduction in maternal employment (exclusion restriction) (Dustmann and Schönberg, 2012). In the concrete example, you divide the ITT estimates by $0.45 \times 0.835 = 0.377$ (share mothers taking ML $\times$ reduction in maternal labor supply) to obtain the effect of spending one additional month away from work after childbirth on children's health outcomes.

[32] The dependent variables are defined on the level of the MOB cohort and aggregated to the area of the former FRG.

Table 1.1. ITT effects on hospital admission

	Estimation window				
	(1) 6M	(2) 5M	(3) 4M	(4) 3M	(5) Donut
Panel A. Over entire length of the life-course					
Overall	-2.076**	-1.872*	-2.176*	-2.214	-2.576***
	(0.772)	(0.905)	(1.126)	(1.399)	(0.813)
Dependent mean	121.1	121.0	121.5	123.3	121.9
Effect in SDs [%]	18.88	16.67	18.77	19.21	23.29
N (MOB × year)	456	380	304	228	380
Panel B. Age brackets					
Age 17-21	-1.517	-0.590	-0.735	-1.095	-1.963**
	(0.946)	(0.995)	(1.198)	(1.603)	(0.931)
Age 22-26	-0.611	-0.613	-0.667	-0.735	-1.080
	(0.937)	(1.113)	(1.412)	(1.672)	(1.012)
Age 27-31	-2.665***	-3.015***	-3.209**	-2.546*	-2.974***
	(0.826)	(0.909)	(1.132)	(1.375)	(0.949)
Age 32-35	-3.869***	-3.619**	-4.572***	-5.045**	-4.717***
	(1.083)	(1.277)	(1.460)	(1.721)	(1.191)

Notes: The table shows DiD estimates of the 1979 maternity leave reform on hospital admission for different estimation windows around the cutoff. The *'Donut'* specification uses a bandwidth of half a year and excludes children born in April and May. Panel A shows the effect for the entire pooled time frame and panel B reports estimates per age bracket. The outcome variables are defined as the number of cases per thousand individuals. All regressions control for year and month-of-birth fixed effects. The control group is comprised of children that are born in the same months but one year before the reform (i.e. children born between November 1977 and October 1978). In order to compare the two birth cohorts at the same age, I shift the control cohort from wave t to wave $t+1$. The dependent mean and the effect size in standard deviation units correspond to pre-reform values of the treated group. Table A.4 contains the number of observations for the estimations per age bracket. Clustered standard errors are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

individuals. This corresponds to a reduction of 1.7% from the pre-treatment mean. The point estimate is robust to using narrower estimation windows, as shown in columns 2 - 4, and to excluding children born close to the cutoff, as presented in column 5. The results obtained when using narrower bandwidths are less precisely estimated due to the smaller sample sizes. However, the point estimates do not differ significantly across the specifications. Panel B contains DiD estimates when dividing the pooled sample into four age brackets and estimating the model for each group separately. Although all estimates are negative, the coefficients for the age cohorts 17-21 as well as 22-26 are small and not significantly different from zero. The average effect on hospitalizations for the age groups 27-31 and 32-35 on the other hand, is large and significant. Consequently, the results indicate that the reform's impact on hospitalization rates is increasing with children's age.

In the next step, I explore whether these general findings hold similarly for men and women. Table 1.2 shows the effect of the 1979 ML reform on hospital admissions for women and men, respectively. In general, the point estimates for men are larger than those for

Table 1.2. ITT effects on hospital admission, by gender

	Women					Men				
	(1) 6M	(2) 5M	(3) 4M	(4) 3M	(5) Donut	(6) 6M	(7) 5M	(8) 4M	(9) 3M	(10) Donut
Panel A. Over entire length of the life-course										
Overall	-1.742**	-1.224	-0.689	-0.862	-2.164***	-2.410**	-2.502*	-3.593**	-3.506**	-2.986**
	(0.816)	(0.924)	(1.117)	(1.504)	(0.718)	(1.015)	(1.204)	(1.373)	(1.568)	(1.178)
Dependent mean	122.3	121.9	121.9	123.8	123.2	120.0	120.2	121.2	122.7	120.7
Effect in SDs [%]	15.30	10.61	5.750	7.350	18.84	19.31	19.68	27.68	26.59	23.72
N (MOB × year)	456	380	304	228	380	456	380	304	228	380
Panel B. Age brackets										
Age 17-21	-2.916***	-1.931*	-1.274	-2.121	-3.322***	-0.273	0.634	-0.246	-0.157	-0.757
	(0.935)	(0.959)	(1.018)	(1.269)	(0.985)	(1.201)	(1.344)	(1.592)	(2.147)	(1.241)
Age 22-26	0.0274	0.557	1.126	0.707	-0.510	-1.230	-1.738	-2.373	-2.113	-1.633
	(1.267)	(1.461)	(1.806)	(2.395)	(1.117)	(1.048)	(1.226)	(1.497)	(1.519)	(1.241)
Age 27-31	-2.762**	-2.605**	-1.669	-1.379	-2.944***	-2.558*	-3.408**	-4.669**	-3.650**	-2.987*
	(1.004)	(1.163)	(1.336)	(1.765)	(0.917)	(1.294)	(1.433)	(1.625)	(1.467)	(1.528)
Age 32-35	-1.212	-0.841	-1.004	-0.605	-1.810*	-6.373***	-6.244***	-7.955***	-9.253***	-7.461***
	(0.866)	(1.024)	(1.165)	(1.300)	(0.941)	(1.526)	(1.781)	(1.969)	(2.318)	(1.722)

Notes: The table shows DiD estimates of the 1979 maternity leave reform on hospital admission by gender. The *'Donut'* specification uses a bandwidth of half a year and excludes children born in April and May. Panel A shows the effect for the entire pooled time frame and panel B breaks the life-course up in age brackets. The outcome variables are defined as the number of cases per thousand individuals. All regressions control for year and month-of-birth fixed effects. The control group is comprised of children that are born in the same months but one year before the reform (i.e. children born between November 1977 and October 1978). In order to compare the two birth cohorts at the same age, I shift the control cohort from wave t to wave $t+1$. The dependent mean and the effect size in standard deviation units correspond to pre-reform values of the treated group. Table A.4 contains the number of observations for the estimations per age bracket. Clustered standard errors are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

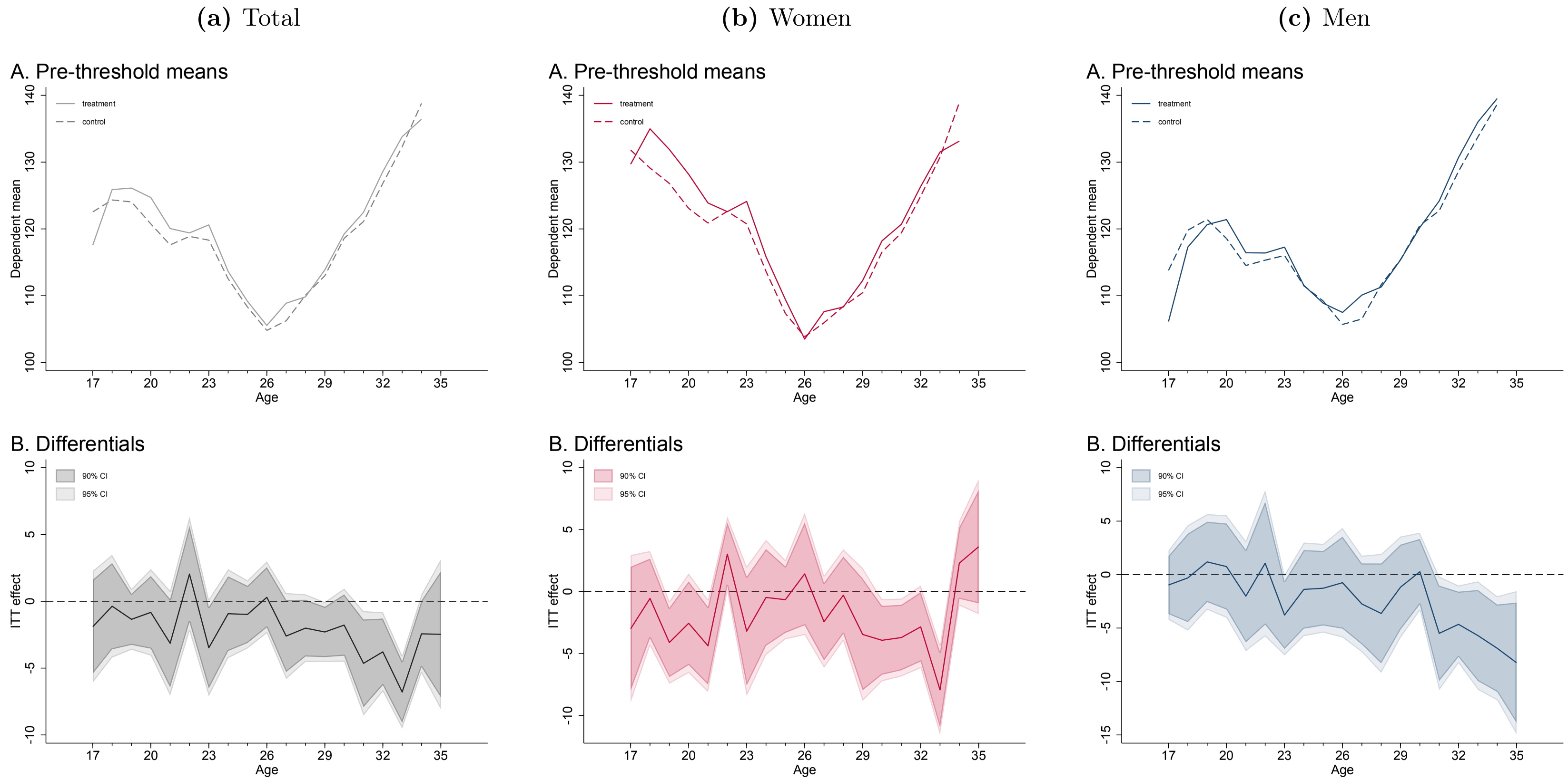


Figure 1.2. Life-course approach for hospital admission

Notes: The top panels show pre-threshold (born November-April) means for the treatment (1978/79, solid line) and control cohort (1977/78, dashed line) across the years. The bottom panels plot DiD estimates (along with 90% and 95% confidence intervals) for the impact of the reform on hospital admissions over the life-course. For a given reporting year ($N = 24$), I estimate the model in equation 1.1 (without ρ_t) and plot the DiD estimate and the corresponding confidence interval for that year. The outcomes are defined as the number of cases per 1,000 individuals. Column a shows the results for all admissions, whereas columns b and c show the estimates for females and males, respectively.

women.[33] Furthermore, the effects for women are less robust to the choice of bandwidth than the effects for men, which mirror the overall effects from Table 1.1 closely. The DiD estimates suggest that men, whose mothers were eligible for extended leave duration, have lower hospitalization rates, irrespective of the estimation window. In the pooled sample, I find an average reduction of 2.4 fewer hospital admissions per 1,000 individuals. Moreover, the effects for males increase in size the older men become.[34]

A finer granularity with respect to age helps to identify the periods when the effects of the reform are particularly salient. The top panels of Figure 1.2 show year-by-year pre-threshold means (born November-April) for the treatment (1978/79) and control (1977/78) cohort (ranging from age 17 up to 35). Importantly for identification, they demonstrate that hospitalization rates for children born before the cutoff in the treatment and control year are parallel over the years. The bottom panels of Figure 1.2 display the trajectories of yearly hospitalization differentials over the same time period.[35] I estimate the specification in equation 1.1 for each year and plot the DiD estimates across time. Column a) shows the effect on all hospital admissions over the life-course. Overall, the average reform effect is close to zero for younger ages, but grows in magnitude and becomes significantly different from zero for older ages. This negative trend, which reflects a growing positive health effect of the reform, is more pronounced for men (column c): The positive health impact of the reform seems to increase with age. In contrast, the figure for women (column b) does not reveal a similar and clear picture: While there are several significant reductions in hospitalization rates for older ages, there is no significant effect of the reform at age 35.

1.5.2 Diagnosis Chapters

What is driving the significant reductions in hospitalization rates? I exploit the detailed reporting on the main diagnosis related to each hospitalization case in the data and assess the effect of extended ML on the components of hospitalizations. For clarification, the dependent variables now refer to the number of specific diagnoses, grouped in 13 chapters, per 1,000 individuals.[36]

Figure 1.3 plots DiD estimates of the policy effect on the incidence rate per diagnosis chapter for all patients (panel A) as well as for women and men separately (panels B and C). The estimates are based on the pooled sample and a bandwidth of six months. In addition to the estimates, each panel shows the frequency distribution across chapters from 1995 to 2014. The results in panel A (all hospitalizations, irrespective of gender) illustrate that almost all point estimates are either negative or not significantly different from zero. This implies that it is unlikely the expansion has a detrimental effect on any long-run child

[33] Yet, the baseline means are not significantly different from each other.

[34] Table A.5 contains a robustness check when using the full sample and interactions of Treat × After with the age brackets. The interaction of being born after the threshold in the treatment year with the age groups addresses potential serial correlation of the errors for a given month of birth cohort. Compared to the baseline results of having different regressions for each age group, the two key insights remain the same. The largest effects are observed for men and in the oldest age bracket. In contrast to the baseline effects, men show significant reductions in hospitalizations in the youngest age bracket, and women have lower hospitalization rates in the oldest age bracket.

[35] I still control for any maturity effects and compare outcomes of treatment and control cohorts measured at the same age. Year fixed effects ρ_t have to be omitted from the specification due to collinearity.

[36] Panel A of Table A.2 presents an overview of the diagnosis chapters.

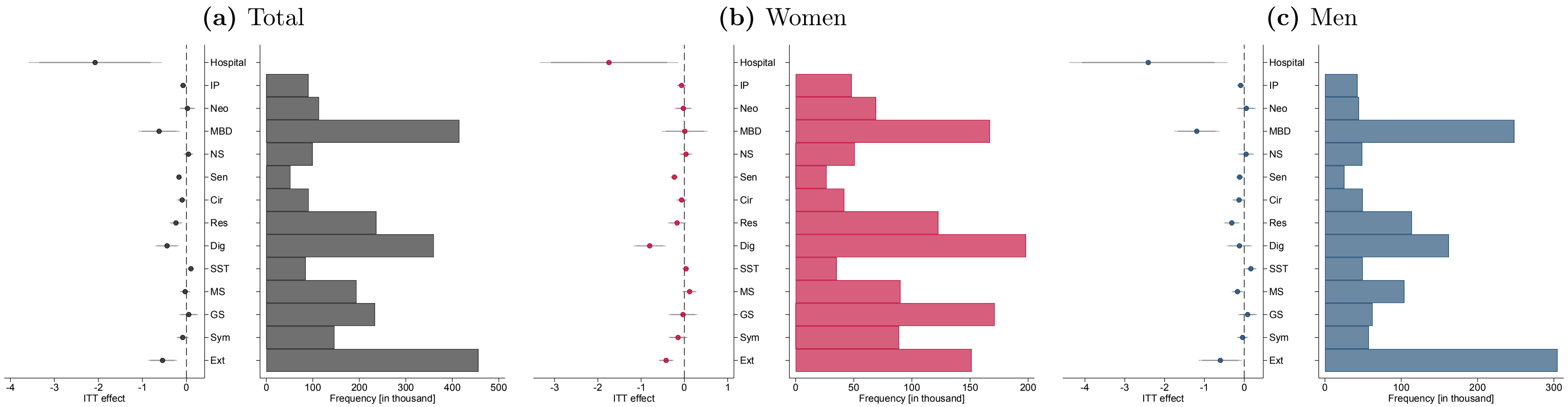


Figure 1.3. Intention-to-treat effects across main diagnosis chapters

Notes: The figures plot intention-to-treat estimates (along with 90%/95% confidence intervals) across the main diagnosis chapters. Furthermore, they indicate how often each chapter is diagnosed over the entire time frame (1995-2014). The outcomes are defined as the number of cases per 1,000 individuals. The point estimates are coming from a DiD regression as described in section 1.4, with a bandwidth of six months, month-of-birth and year fixed effects, and clustered standard errors on the month-of-birth level. The control group is comprised of children that are born in the same months but one year before the reform (i.e. children born between November 1977 and October 1978).

Legend: Infectious and parasitic diseases (IPD), neoplasms (Neo), mental and behavioral disorders (MBD), diseases of the nervous system (Ner), diseases of the sense organs (Sen), diseases of the circulatory system (Cir), diseases of the respiratory system (Res), diseases of the digestive system (Dig), diseases of the skin and subcutaneous tissue (SST), diseases of the musculoskeletal system (Mus), diseases of the genitourinary system (Gen), symptoms, signs, and ill-defined conditions (Sym), injury, poisoning and certain other consequences of external causes (Ext).

Table 1.3. ITT effects on hospital admission and main diagnoses chapters

	(1)	(2)	(3)	(4)	(5)
	Overall	Age brackets [years]			
		17-21	22-26	27-31	32-35
Hospital	-2.076**	-1.517	-0.611	-2.665***	-3.869***
	(0.772)	(0.946)	(0.937)	(0.826)	(1.083)
IPD	-0.0763*	0.0130	-0.162	-0.126**	-0.0197
	(0.0379)	(0.0680)	(0.108)	(0.0548)	(0.111)
Neo	0.0221	-0.198	0.336**	0.121	-0.217
	(0.0917)	(0.155)	(0.139)	(0.102)	(0.164)
MBD	-0.621**	0.174	-0.00769	-1.000**	-1.906***
	(0.242)	(0.263)	(0.420)	(0.357)	(0.372)
Ner	0.0444	-0.0919	0.238***	0.0374	-0.0190
	(0.0564)	(0.0558)	(0.0751)	(0.0963)	(0.126)
Sen	-0.170***	-0.168**	-0.0990*	-0.172**	-0.261**
	(0.0292)	(0.0626)	(0.0518)	(0.0640)	(0.0998)
Cir	-0.0981	-0.0466	0.0311	-0.199*	-0.198
	(0.0632)	(0.0782)	(0.0812)	(0.0994)	(0.162)
Res	-0.239***	-0.369**	-0.199***	-0.273**	-0.0842
	(0.0736)	(0.173)	(0.0694)	(0.107)	(0.167)
Dig	-0.441***	-0.421**	-0.485*	-0.376	-0.494
	(0.137)	(0.177)	(0.257)	(0.227)	(0.361)
SST	0.101**	0.0567	0.186**	0.152*	-0.0149
	(0.0385)	(0.0947)	(0.0721)	(0.0882)	(0.120)
Mus	-0.0304	-0.0260	-0.155	-0.0562	0.152
	(0.0619)	(0.0884)	(0.172)	(0.141)	(0.158)
Gen	0.0524	0.241	0.180	-0.194	-0.0356
	(0.111)	(0.159)	(0.207)	(0.235)	(0.159)
Sym	-0.0853	-0.175	0.112	-0.164**	-0.122
	(0.0704)	(0.151)	(0.0988)	(0.0604)	(0.103)
Ext	-0.543***	-0.460	-0.685***	-0.429**	-0.612***
	(0.168)	(0.328)	(0.230)	(0.193)	(0.135)
N (MOB × year)	456	120	120	120	96

Notes: This table reports DiD estimates across the main diagnosis chapters for the entire life-course or per age bracket. Each row corresponds to a different estimation with the number of diagnoses per 1,000 individuals. See Table 1.1 for additional details. Clustered standard errors are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Legend: Infectious and parasitic diseases (IPD), neoplasms (Neo), mental and behavioral disorders (MBD), diseases of the nervous system (Ner), diseases of the sense organs (Sen), diseases of the circulatory system (Cir), diseases of the respiratory system (Res), diseases of the digestive system (Dig), diseases of the skin and subcutaneous tissue (SST), diseases of the musculoskeletal system (Mus), diseases of the genitourinary system (Gen), symptoms, signs, and ill-defined conditions (Sym), injury, poisoning and certain other consequences of external causes (Ext).

health outcome, as represented by the chapters in use.[37] Furthermore, the ML extension has the largest impact on MBDs, followed by consequences of external causes (injuries), diseases of the digestive system, and respiratory maladies. Comparing the results against the frequency of diagnoses chapters, the strong effect on MBDs is striking as this diagnosis chapter has a particularly high prevalence rate among this age group. When looking at heterogeneous effects across genders, the largest impact for both women and men occur for the chapter that is among the most frequently diagnosed. For women, the largest reduction in hospitalizations is the result of fewer diagnoses of digestive diseases, whereas for males the reduction stems from fewer diagnoses of MBDs.

Breaking up the analysis by age group generates additional insights. Table 1.3 contains DiD estimates of the impact of the ML expansion on the main diagnosis chapters for all inpatients. Column 1 reports the effect on the pooled sample and is equivalent to the estimates in the graphical representation. Columns 2 to 5 list the impact of the expansion on the main diagnosis chapter per age bracket.[38] Some chapters exhibit increasing health differentials as individuals grow older. This is especially the case for MBDs, which resemble the overall effects of hospital admissions. In general, MBDs account for the largest relative contribution in hospitalization reduction. In the pooled sample, MBDs account for one-third of the reduction in hospitalizations, and in the last age bracket (32-35 years of age) the importance of this diagnosis chapter for the drop in hospitalizations rises to almost 50%. The effects on injuries are responsible for almost one-fourth of the reduction in hospital admissions and are mostly constant across age groups. The estimates indicate that there are, on average, between 0.4 and 0.7 fewer injuries per 1,000 individuals across age groups. Other chapters display differentials that fade out with increasing age, such as diseases of the digestive system (around 0.4 fewer diagnoses per 1,000 individuals, which corresponds to 20 percent of the decline in hospitalizations) and respiratory maladies (0.2-0.4 fewer diagnoses per 1,000 individuals; $\sim$ 11 percent of the decline in hospitalizations).

1.5.3 Mental and Behavioral Disorders

MBDs stand out in terms of importance for the overall effect of the ML reform on hospitalizations and its prevalence in the age group under consideration. For this reason, Table 1.4 presents more refined evidence by reporting DiD estimates for the effect of the ML expansion on the diagnosis of MBDs for all inpatients. The coefficient of the pooled sample in Panel A, column 1 suggests that over the entire time frame, there are on average 0.6 fewer diagnoses of MBDs per 1,000 individuals. This corresponds to a decline of 3.2% from the pre-treatment mean. When looking at the effects per age bracket, as shown in Panel B, no significant effects for younger cohorts and an increasing impact of the ML expansion with children's age are observed. Table 1.5 shows the effects for women and men separately. While the estimates for women are close to zero and not statistically significant from zero, the magnitude of the effects for men is large. In the pooled sample, I find an average reduction of 1.2 MBD diagnoses per 1,000 individuals for men who were born after the ML expansion. When considering the estimates per age bracket, there is a similar pattern to hospitalization rates: Health differentials open up from the age of 27

[37] Diseases of the skin and subcutaneous tissue are a notable exception. Their positive DiD estimate is, albeit small in magnitude, statistically significant at the 5 percent level. As such, the unexpected positive coefficient might be a statistical aberration, but it contravenes my hypothesis of favorable health effects.

[38] Appendix Figure A.8 shows the respective life-course figures for each diagnosis chapter.

Table 1.4. ITT effects on mental & behavioral disorders

	Estimation window				
	(1) 6M	(2) 5M	(3) 4M	(4) 3M	(5) Donut
Panel A. Over entire length of the life-course					
Overall	-0.621**	-0.734**	-0.853**	-0.688	-0.789***
	(0.242)	(0.272)	(0.336)	(0.423)	(0.262)
Dependent mean	19.57	19.59	19.67	19.84	19.77
Effect in SDs [%]	12.44	14.39	16.43	13.07	15.92
N (MOB $\times$ year)	456	380	304	228	380
Panel B. Age brackets					
Age 17-21	0.174	0.268	0.318	0.135	-0.0603
	(0.263)	(0.314)	(0.387)	(0.516)	(0.239)
Age 22-26	-0.00769	-0.146	-0.172	0.343	-0.360
	(0.420)	(0.500)	(0.607)	(0.640)	(0.454)
Age 27-31	-1.000**	-1.301***	-1.508***	-1.258**	-1.020**
	(0.357)	(0.391)	(0.478)	(0.546)	(0.433)
Age 32-35	-1.906***	-2.015***	-2.352***	-2.293***	-1.949***
	(0.372)	(0.295)	(0.305)	(0.365)	(0.439)

Notes: The table shows DiD estimates of the 1979 maternity leave reform on mental and behavioral disorders for different estimation windows around the cutoff. The *'Donut'* specification uses a bandwidth of half a year and excludes children born in April and May. Panel A shows the effect for the entire pooled time frame and panel B breaks the life-course up in age brackets. The outcome variables are defined as the number of cases per thousand individuals. All regressions control for year and month-of-birth fixed effects. The control group is comprised of children that are born in the same months but one year before the reform (i.e. children born between November 1977 and October 1978). In order to compare the two birth cohorts at the same age, I shift the control cohort from wave t to wave $t+1$. The dependent mean and the effect size in standard deviation units correspond to pre-reform values of the treated group. Table A.4 contains the number of observations for the estimations per age bracket. Clustered standard errors are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and further increase towards the end of the observed time span. The bottom panels of Figure 1.4 shows the corresponding life-course graphs, which confirm the findings obtained from the tables. Column a) shows DiD estimates on a yearly basis for all diagnoses related to MBD. The life-course graph illustrates that at younger ages the effect of the legislation change is close to zero and insignificant. Yet, the estimates grow in magnitude and become statistically significant for older cohorts. Overall, the reduction in MBD diagnoses at older ages appear to be driven by males. Their life-course graph exhibits even more pronounced features than the corresponding one for all cases in panel a). In contrast, the differentials for women do not appear to follow a visible trend. The top panels in Figure 1.4 show that the parallel trends assumption holds as the difference in pre-threshold means of treatment and control cohort is reasonably constant over time.

Next, I investigate the drivers for the reduction in MBDs in response to the 1979 ML reform. To do so, I make use of the detailed coding in the ICD classification system and investigate the effect on diseases that constitute the chapter of mental and behavioral

Table 1.5. ITT effects on mental & behavioral disorders, by gender

	Women					Men				
	(1) 6M	(2) 5M	(3) 4M	(4) 3M	(5) Donut	(6) 6M	(7) 5M	(8) 4M	(9) 3M	(10) Donut
Panel A.Over entire length of the life-course										
Overall	0.00972	-0.0900	-0.205	-0.244	0.0211	-1.192***	-1.328***	-1.462***	-1.098**	-1.533***
	(0.271)	(0.303)	(0.377)	(0.496)	(0.285)	(0.288)	(0.336)	(0.412)	(0.486)	(0.286)
Dependent mean	16.11	16.09	16.09	16.32	16.32	22.84	22.91	23.07	23.19	23.05
Effect in SDs [%]	0.300	2.650	5.940	7.010	0.650	17.60	19.29	20.84	15.43	22.64
N (MOB $\times$ year)	456	380	304	228	380	456	380	304	228	380
Panel B. Age brackets										
Age 17-21	0.388	0.416	0.527	0.0745	0.235	-0.0319	0.129	0.119	0.192	-0.344
	(0.313)	(0.378)	(0.463)	(0.555)	(0.318)	(0.262)	(0.300)	(0.373)	(0.507)	(0.217)
Age 22-26	0.205	0.119	0.0485	0.217	-0.0753	-0.180	-0.379	-0.373	0.475	-0.602
	(0.466)	(0.558)	(0.700)	(0.791)	(0.499)	(0.485)	(0.575)	(0.683)	(0.680)	(0.526)
Age 27-31	-0.426	-0.598	-0.816	-0.612	-0.273	-1.504***	-1.943***	-2.152***	-1.854**	-1.690***
	(0.418)	(0.469)	(0.579)	(0.781)	(0.426)	(0.507)	(0.542)	(0.675)	(0.652)	(0.570)
Age 32-35	-0.163	-0.349	-0.671*	-0.760	0.242	-3.518***	-3.568***	-3.938***	-3.733***	-3.989***
	(0.388)	(0.335)	(0.344)	(0.466)	(0.396)	(0.515)	(0.522)	(0.596)	(0.741)	(0.515)

Notes: The table shows DiD estimates of the effect of the 1979 maternity leave reform on mental and behavioral disorders for different estimation windows around the cutoff. The *'Donut'* specification uses a bandwidth of half a year and excludes children born in April and May. Panel A shows the effect for the entire pooled time frame and panel B breaks the life-course up in age brackets. The outcome variables are defined as the number of cases per thousand individuals. All regressions control for year and month-of-birth fixed effects. The control group is comprised of children that are born in the same months but one year before the reform (i.e. children born between November 1977 and October 1978). In order to compare the two birth cohorts at the same age, I shift the control cohort from wave t to wave $t+1$. The dependent mean and the effect size in standard deviation units correspond to pre-reform values of the treated group. Table A.4 contains the number of observations for the estimations per age bracket. Clustered standard errors are reported in parentheses.
Significance levels: * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

(a) Total **(b)** Women **(c)** Men

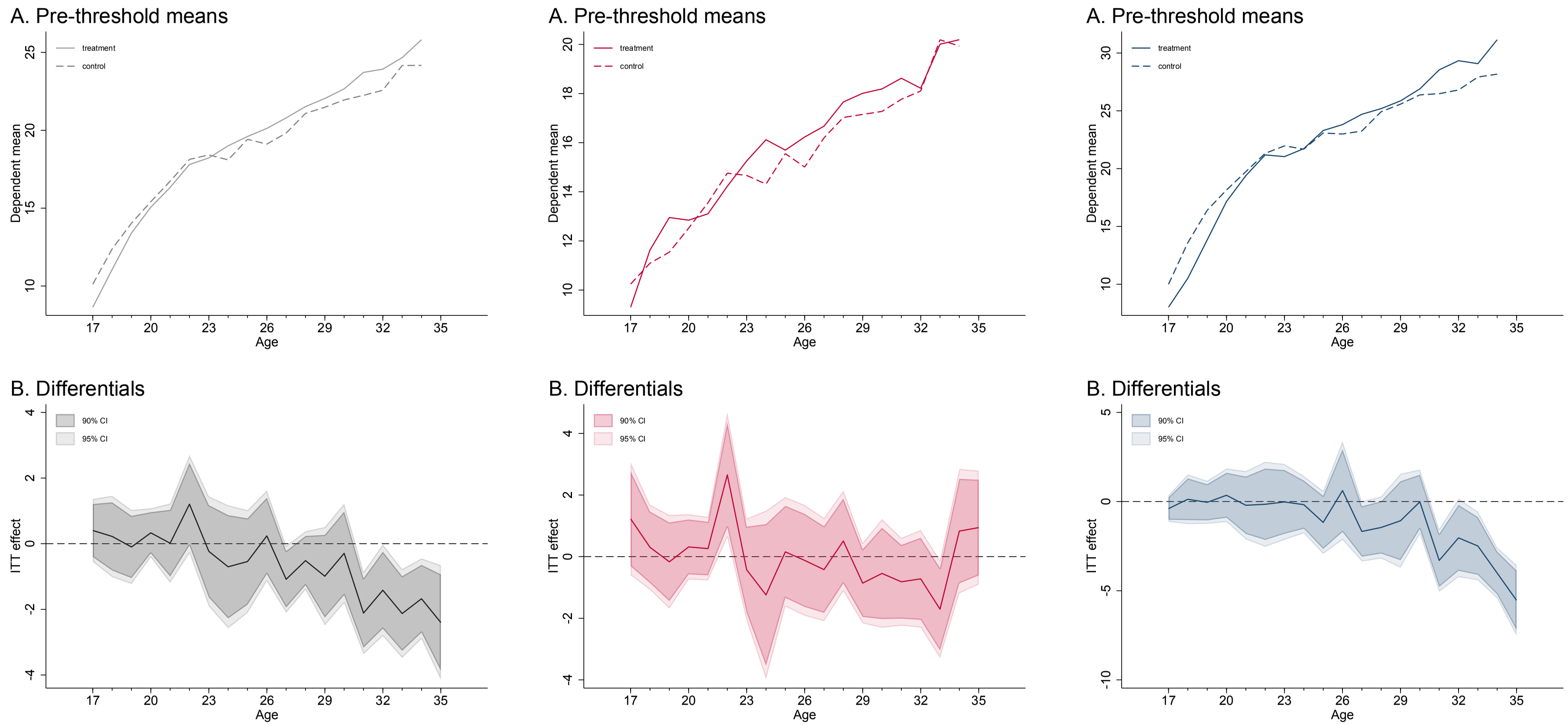


Figure 1.4. Life-course approach for mental and behavioral disorders

Notes: The top panels show pre-threshold (born November-April) means for the treatment (1978/79, solid line) and control cohort (1977/78, dashed line) across the years. The bottom panels plot DiD estimates (along with 90% and 95% confidence intervals) for the impact of the reform on mental and behavioral disorders over the life-course. For a given reporting year ($N = 24$), I estimate the model in equation 1.1 (without ρ_t) and plot the DiD estimate and the corresponding confidence interval for that year. The outcomes are defined as the number of cases per 1,000 individuals. Column a shows the results for all admissions, whereas columns b and c show the estimates for females and males, respectively.

(a) Total **(b)** Women **(c)** Men

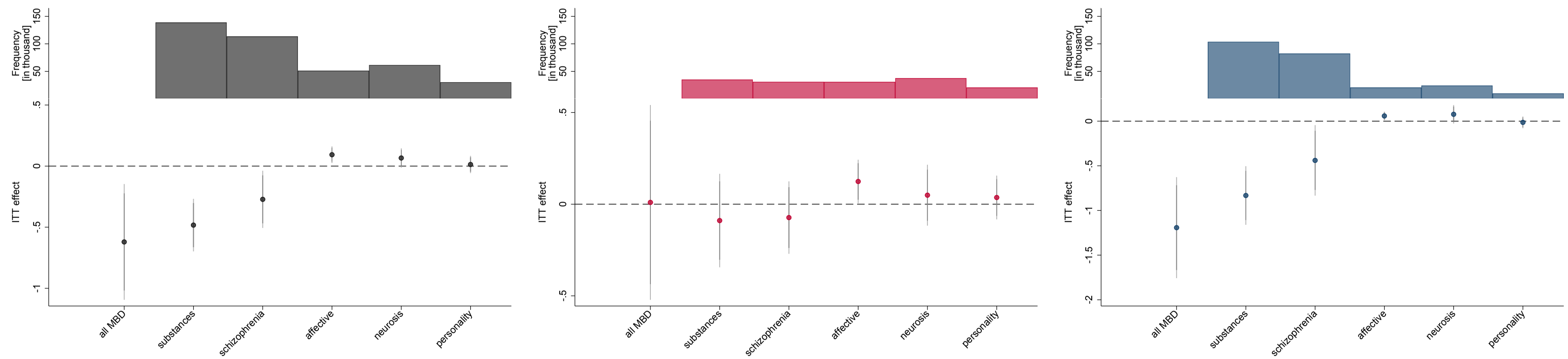


Figure 1.5. ITT effect for subcategories of mental and behavioral disorders
Notes: The figure plots ITT estimates (along with 90%/95% confidence intervals) across the five most common subcategories of MBDs. Moreover, they indicate how often each subcategory is diagnosed over the time window of 1995-2014. The outcomes are defined as the number of cases per 1,000 individuals. The point estimates are coming from a DiD regression as described in section 1.4, with a bandwidth of six months, month-of-birth and year fixed effects, and standard errors clustered at the month-of-birth level. The control group is comprised of children that are born in the same months but one year before the reform (i.e. children born between November 1977 and October 1978).

diseases (henceforth called subdiagnoses or subcategories).[39] Appendix Figure A.6 shows the temporal variation in the composition of mental and behavioral diseases. In 2014, the endpoint of the sample, the five most frequent MBDs are (in descending order): mental and behavioral disorders due to the use of psychoactive substances, schizophrenia, affective disorders, neuroses, and finally personality disorders. These five subcategories constitute more than 95% of all MBDs. Figure 1.5 shows DiD estimates of the legislation effect and incidence rates of the subcategories of MBDs.[40] Panel A shows that the reduction in MBDs is the result of fewer diagnoses related to the use of psychoactive substances and fewer incidences of schizophrenia. When considering effect heterogeneity by gender, it can be seen that the reduction is driven by fewer diagnoses for males.

1.5.4 Robustness Tests

I perform several sensitivity and placebo tests to assess the robustness of the findings. The results of these checks are reported in Table 1.6. Overall, the sensitivity tests demonstrate that the main results are robust to alternative specifications and estimations, indicating that the ML reform significantly reduces hospitalization rates between the ages 17 to 35, and particularly for men.

Alternative Specifications.—First, I test whether the results are sensitive to an alternative specification of the *dependent variable*, namely the number of diagnoses per 1,000 individuals born in a specific month-year combination. Due to the lack of yearly population figures on the level of month-of-birth in a given year-of-birth, I thus resort in the main analysis to the number of births per birth month as denominator, which is time-invariant. However, over time, this denominator might reflect the actual population size of a given MOB cohort imperfectly (due to migration or death). Therefore, I construct an alternative denominator using actual annual population figures by year of birth, weighted by the relative frequency of births across birth months in that particular year of birth.[41] Column 2 in Table 1.6 shows that my results are robust to the use of this alternative denominator.[42] Second, I show that the results are not sensitive to a different *level of aggregation.* For this exercise, I spatially disaggregate the data and create a regional panel for 204 labor market regions (LMR) in West Germany covering the years 2003 - 2014. Figure A.7 contains a map with the LMR used in the analysis. The dependent variable is using the same denominator specification as in the preceding robustness test, but on the regional level of the LMR. In order to avoid potentially confounding effects, I include region fixed effects.[43] Column 3

[39]For instance, the chapter of mental and behavioral diseases has codes starting with F00-F99. In this exercise I go one level deeper and look at the effect of the reform on, for example, affective disorders, which are grouped in the codes F30-F39. Panel B of Table A.2 in the Appendix defines the subcategories of the mental and behavioral diseases chapter.

[40]Appendix Table A.6 presents results when dividing the pooled sample into four age brackets. The reductions in the subcategories are experienced by the age cohorts 27-31 and 32-35, and tend to get larger with increasing age.

[41]For instance, to get the number of people born in May 1979 who live in Germany in 2014, I use the number of people born in 1979 (observed in 2014) and weigh them with the fraction $\frac{\text{births in May 1979}}{\text{births in 1979}}$.

[42]The number of observations is smaller since the statistics on annual population size by birth month and year are only available since 2003.

[43]The corresponding regression specification is as follows (regions are weighted by population):
$\mathrm{Y}_{mrt} = \gamma_0 + \gamma_1 \mathrm{Treat}_m + \gamma_2 \mathrm{After}_m + \gamma_3(\mathrm{Treat}_m \times \mathrm{After}_m) + \psi_m + \phi_r + \rho_t + \varepsilon_{mrt}.$

Table 1.6. Robustness checks for hospital admission

		Alternative specifications		Alternative estimation		Placebos	
	(1) Baseline	(2) Current population	(3) LMR level[a]	(4) DDD[b]	(5) Add. CG	(6) Temporal: cohort	(7) Spatial: GDR
(1) Total	-2.076**	-1.581**	-1.771***	-2.313*	-2.293**	-0.203	0.154
	(0.772)	(0.675)	(0.623)	(1.127)	(0.987)	(0.962)	(0.469)
(2) Female	-1.742**	-0.694	-0.740	-1.255	-1.559	0.561	-0.396
	(0.816)	(0.633)	(0.597)	(1.231)	(1.112)	(0.964)	(0.503)
(3) Male	-2.410**	-2.462**	-2.816***	-3.252**	-3.007**	-0.926	0.593
	(1.015)	(0.981)	(0.945)	(1.310)	(1.135)	(1.081)	(0.714)
For total:							
Dependent mean	121.1	92.22	98.66	121.8	121.1	120.2	66.29
Effect in SDs [%]	18.88	16.21	4.750	20.94	20.86	1.900	1.260
N	456	288	53,855	912	672	456	456
MOB fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: This table displays robustness checks for the effect of the 1979 maternity leave reform on hospital admissions. I perform the following checks (with reference to the column): (1) baseline specification that was used in previous parts of the paper, (2) for the outcome I use the number of diagnoses divided by the current number of individuals (approximation), (3) the analysis is carried out on the level of labor market regions, (4) triple difference model (the third difference stems from the former region of the GDR), (5) I use as control cohort not only the cohort before the reform, but also the cohort 2 years prior to the policy change, (6) first placebo, in which the entire analysis set-up is pushed back by one year, i.e. the placebo TG is the cohort prior to the real TG and the placebo CG is the cohort born 2 years before the reform took place, (7) second placebo, in which I run the normal DiD set-up in the area of the former GDR.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[a]: level of analysis on Labor Market Regions: weighted regressions (by population), includes region fixed effects.
[b]: standard errors clustered on the month-of-birth×birth-cohort×East-West cell level.

presents estimates based on the regional panel data that are similar to the main results.

Alternative Estimations.—In the next step, I investigate the effect of changes in the estimation approach. First, column 4 of Table 1.6 reports estimates based on a *triple-differences (DDD)* approach. The additional control group is comprised of individuals living on the territory of the former GDR (East Germany). The underlying idea is that children (and their mothers) born in East Germany in 1979 were not affected by the West German 1979 ML reform and should consequently resemble a valid comparison group, conditional on the common trend assumption. Importantly, this specification allows me to net out general time trends in the outcome variables that might affect the results. The corresponding triple-differences model is as follows:

$$\begin{aligned} \mathrm{Y}_{mt} = \beta_0 &+ \beta_1 \mathrm{Treat}_m + \beta_2 \mathrm{After}_m + \beta_3 \mathrm{FRG}_m \\ &+ \beta_4(\mathrm{Treat}_m \times \mathrm{After}_m) + \beta_5(\mathrm{Treat}_m \times \mathrm{FRG}_m) + \beta_6(\mathrm{After}_m \times \mathrm{FRG}_m) \\ &+ \beta_7(\mathrm{Treat}_m \times \mathrm{After}_m \times \mathrm{FRG}_m) + \psi_m + \rho_t + \varepsilon_{mt} \end{aligned} \tag{1.2}$$

which now contains an additional dummy variable FRG_m (West Germany) as well as interactions of this group with the treatment cohort 1979 (Treat_m), children born in/after May (After_m) and with the interaction term $\mathrm{Treat}_m \times \mathrm{After}_m$. The parameter of interest is β_7, which captures the impact of the 1979 ML reform on health outcomes. Likewise, the general pattern of the main results remains robust to this alternative estimation. While the estimates become less precise, the point estimates increase slightly but not significantly.

Second, as a further sensitivity test, I re-estimate the main regression equation including an *additional control group* next to the one defined in section 1.4. The additional control group is comprised of children born in the months around the threshold month of May, but two years before the reform took place (i.e. around May 1977). The advantage of including another pre-reform West German birth cohort is that it may account better for systematic month-of-birth patterns in hospitalization rates due to a larger sample size. The estimates from this approach can be found in column 5. Once again, the results and pattern of the baseline results hold when adding this additional control cohort.

A potential concern is that the 1979 ML reform had positive spillover effects on siblings of treated children, which may be motivated by more time with the mother, better maternal health, among other reasons (see discussion next section). Positive spillover effects would imply a violation of the *stable unit treatment value assumption* (SUTVA) and the obtained DiD estimates would represent a lower bound of the true effect. The baseline specification uses the cohort born one year prior to the treatment cohort. On the one hand, this helps to limit the possibility of spillover effects as they would only be present in case of very short birth spacing. On the other hand, the close temporal proximity to the treatment cohort may be problematic, as siblings close in age will benefit more from the ML expansion than older siblings. The previous robustness check addresses this issue by including another older control cohort. Figure A.10 goes one step further and presents suggestive evidence supporting SUTVA. If spillovers were present, one would expect the estimates to vary in magnitude, dependent on which control group is used. However, the DiD estimates of the effect of the 1979 ML reform on hospitalizations are fairly stable and reflect the results from the baseline specification well, irrespective of whether the control group is composed of the cohort born one, two, three, or four years before the treatment group. The same is true for the estimates by gender.

Placebo Tests.—I perform two placebo tests to validate my identifying assumptions: First, I run a *temporal placebo* and test whether the estimated effects are caused by children eligible to the more generous ML regime and not by an underlying general time trend affecting children born post-April. For this exercise, I use the two pre-reform birth cohorts 1978 and 1977, using May 1, 1978 as the placebo reform date. The estimates from this placebo analysis can be found in column 6. Second, I run a *spatial placebo* analysis and re-estimate the baseline DiD specification but substitute it with the sample of East German individuals who were not affected by the reform. Column 7 shows the estimates of this placebo analysis. In support of the internal validity of the main estimates, both placebo tests yield insignificant effects, further suggesting that the West German ML reform caused a reduction in hospital admissions.

Mental and Behavioral Disorders.—Since this diagnosis chapter plays a central role in this analysis, I run the same set of sensitivity test as for hospital admissions. The results are presented in Table A.7 in the Appendix, and show that the main estimates are robust to the different specifications and estimation procedures.

1.6 Discussion

1.6.1 Conceptual Framework for Long-Run Health Effects

The notion that later-life outcomes originate from early childhood is not novel. In 1990, Barker postulated that conditions in-utero and during infancy have long-lasting effects on later life health. In the study by Shonkoff *et al.* (2009), adult physical and mental well-being was found to be influenced by early experiences, both good and bad, in at least two ways.

On the one hand, early experiences operate via a *cumulative* process, in which physically and psychologically stressful events are experienced repetitively. The persistent experience of stressful events subsequently causes a constant provocation of neurobiological responses, which may precipitate chronic health impairments. Under normal circumstances, however, these responses are healthy and protective because they help to cope with stress.[44] Yet, due to the repetitive activation, the neurobiological reactions may become pathogenic.

On the other hand, the environment at critical developmental stages is *biologically embedded* into regulatory physiological systems such that it can impact adult disease and risk factors latently. During these sensitive periods, the developing brain's architecture is modified considerably and is particularly susceptible to environmental stimuli. The process in which experiences are 'programmed' into the brain's architecture starts in the embryonic state and culminates in the first years of life (Räikkönen *et al.*, 2012). As not all brain circuits develop at the same time, the timing of the experience is crucial.[45] A stimulus has the highest impact on the region of the brain that is undergoing the most changes.[46] During

[44]The neurobiological reactions can include the release of stress hormones, higher blood pressure and heart rate, and the protective mobilization of nutrients, among others. See for example McEwen (1998) and Shonkoff *et al.* (2009).

[45]See the 'life-cycle' model of stress (Lupien *et al.*, 2009).

[46]The effects on later life health impairment are not only determined by the timing but also by the type of the experience (Räikkönen *et al.*, 2012).

infancy, the period of life that is affected by the reform, the hippocampus is the area of the brain that matures most rapidly and is consequently more vulnerable to stimuli than during other stages. This region has been documented to regulate emotions, social behavior, stress responsiveness, and ultimately mental health (Center on the Developing Child at Harvard University, 2016, Shonkoff *et al.*, 2009).

Irrespective of the mechanism at play (the cumulative exposure and the biological embedding), the effects of experiences made in early life may be latent at first until the onset of a particular condition (Almond and Currie, 2011). The time lag could be many years, or even decades (Shonkoff *et al.*, 2009).

1.6.2 Potential Mechanisms

How did the 1979 ML reform affect children's health outcomes in the long-run? After demonstrating that the reform had a higher impact in areas where potentially more women were eligible, I provide a short theoretic discussion on how the 1979 ML reform may have altered the child's environment.

As previously discussed, working mothers postponed their return to work in response to the reform (Dustmann and Schönberg, 2012, Schönberg and Ludsteck, 2014), which allowed more maternal time and family income during a crucial period in a child's development. In the following, I present suggestive evidence for these mechanisms by leveraging heterogeneous eligibility for ML. In particular, I assess whether the reform has a different impact on children's health in rural and urban areas. One reason why I might estimate a differential impact of the reform is that mothers in urban and rural regions differed in their propensity to work. Female labor force participation rates have traditionally been higher in urban areas (Bender and Hirschenauer, 1993), which imply a higher share of eligible mothers in these regions (as only working mothers were eligible for ML). With relatively more children affected by the reform, I expect the ITT estimates to be larger in urban areas. To shed light on maternal return to work behavior as a potential mechanism, I use the regional panel on inpatient cases, as defined by the first set of robustness sets - see column 3 in Table 1.6. Table 1.7 contains DiD estimates for the effect of the ML expansion on health outcomes for inpatients living in rural and urban areas. I label labor market regions as urban if their population density exceeds the median value of all regions.[47] Panel A presents the impact on hospital admissions. Although the estimates have roughly the same size, the standard errors in the urban sample are considerably smaller. Similarly, Panel B shows that urban areas drive the effect on MBDs. The impact of the reform on MBDs in rural areas is not significantly different from zero, whereas the effects in urban areas are large in magnitude and statistically significant.

Taken together, the finding that urban areas drive the overall effects is consistent with the idea that the reform had a larger impact on regions where there were higher female labor force participation rates and hence more women were eligible for ML in 1979. In the next step, I discuss potential mechanisms through which the reform could affect child health outcomes. For this purpose, I consider differences in care quality, parental health differentials, and changes in family income as potential pathways.

[47] Appendix Figure A.7 shows the spatial variation of population density across German labor market regions.

Table 1.7. Subgroup analysis

	Heterogeneity	
	(1) Rural	(2) Urban
Panel A: Hospital admissions		
DiD estimate	-1.654	-1.799***
	(1.096)	(0.598)
Dependent mean	101.3	96.50
Effect in SDs [%]	3.880	5.600
N	24,287	29,568
Panel B: Mental and behavioral disorders		
DiD estimate	-0.241	-0.986***
	(0.564)	(0.196)
Dependent mean	17.00	18.61
Effect in SDs [%]	1.310	7.100
N	24,287	29,568
MOB fixed effects	✓	✓
Year fixed effects	✓	✓
Region fixed effects	✓	✓

Notes: The table shows DiD estimates of the 1979 maternity leave reform on hospital admission for rural and urban areas. The DiD estimates stem from weighted regression (by population) over the entire pooled time frame, and a bandwidth of half a year around the cutoff. The effects in rural and urban areas are not significantly different from each other. The level of analysis is on Labor Market Regions. Figure A.7 shows a map of Germany with the regions marked as rural/urban.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

First, the reform induced mothers to postpone their return to the labor market and allowed more maternal time during a crucial time period of a child's development. With longer ML duration, mothers are more likely to breastfeed and they do so for an extended period (Baker and Milligan, 2008, Albagli and Rau, 2018).[48] The advantages for children that were breastfed range from reduced incidence or severity of asthma, allergies, diarrhea, mortality, morbidity and chronic conditions in the short run, to lower prevalence rates of overweight as well as obesity, and type II diabetes in adulthood (Ruhm, 2000, Victora *et al.*, 2016).[49] Moreover, correlational evidence suggests that the length of breastfeeding is negatively associated with mental health problems and adverse health behavior, such

[48] Baker and Milligan (2008) exploit a ML expansion in Canada, which increased benefit entitlement and job protection from 6 months to a year. Albagli and Rau (2018) investigate an expansion in Chile, in which paid leave was raised from 12 to 24 weeks.

[49] The German Health Interview and Examination Survey for Children and Adolescents (KiGGS) offers representative data on breastfeeding rates from 1986 onwards (Lange *et al.*, 2007). From the West-German 1986 cohort born, around 75% of children were breastfed at least once and the share of children who was breastfed exclusively for half a year is roughly 38%.

as drinking (Oddy *et al.*, 2010, Falk and Kosse, 2016).[50] Apart from these direct health effects, there are also indirect effects of breastfeeding via third outcomes, which in turn may affect health. For example, breastfeeding exhibits a positive effect on cognitive development (Albagli and Rau, 2018), educational attainment and income (Victora *et al.*, 2015), the formation of preferences (Falk and Kosse, 2016), and the quality of mother-child interactions (Papp, 2014). In addition to reduced breastfeeding, early maternal employment impedes the monitoring of children's health status. Berger *et al.* (2005) present associations of early maternal employment and a decrease in the use of preventive health care services (immunizations and 'well-baby' visits), while at the same time problems with externalizing behavior exacerbate. According to Morrill's (2011) instrumental variable estimates, maternal employment leads to a higher likelihood that children suffer from an adverse health event, such as overnight hospitalization, asthma episode, or injury/poisoning.

Second, another mechanism through which the reform might impacted child health outcomes are changes in maternal health outcomes, which could successively affect the ability to nurture.[51] There is correlational evidence indicating that higher maternal employment is related to lower levels of maternal mental well-being as well as self-rated overall health, and a higher frequency of depressive symptoms and problems with parenting stress (Chatterji and Markowitz, 2005, Chatterji *et al.*, 2013). Besides this evidence, there exists a large body of quasi-experimental literature. Beuchert *et al.* (2016) exploit a reform of the parental leave scheme in Denmark and find positive effects on health outcomes of mothers and siblings with more considerable gains for low-resource families. Bütikofer *et al.* (forthcoming) exploit the 1977 ML reform in Norway in order to demonstrate how the legislation change enhances a battery of mid- and long-term maternal health outcomes, such as BMI, blood pressure, pain as well as mental health, and more favorable health behavior, such as physical exercise and smoking abstinence.[52] Albagli and Rau (2018) show that mothers who give birth under a more generous leave regime have lower stress indices as compared to mothers who are shorter on leave. All in all, these studies suggest that extending ML enhances maternal health, which may benefit parents' ability to nurture and ultimately children's health outcomes in adulthood. For instance, maternal postnatal depression is associated with long-term impairment of mother-child bonding (Tronick and Reck, 2009), which the psychological literature links to the development of mental disorders later in life (Canetti *et al.*, 1997).[53] Even the detailed results concerning gender differences and types of affected mental disorders obtained in this study are consistent with the literature. Enns *et al.* (2002) find adult mental health consistently associated with parenting experiences made with one's mother. Although the effects of parenting are diagnostically non-specific, there appear unique effects among men in externalizing disorders, such as substance use and antisocial personality disorder. The generally higher incidence rate of externalizing disorders for males may also explain why the results found in this study are more pronounced

[50] Higher breastfeeding rates may explain next to the decline in MBDs, also the reductions in injuries, digestive diseases, and respiratory illnesses. Falk and Kosse (2016) show that more prolonged breastfeeding is associated with lower willingness to take risks, which may rationalize the findings on injuries. There is also some evidence that breastfeeding may be protective for developing chronic respiratory illnesses such as asthma (Lodge *et al.*, 2015, Friedman and Zeiger, 2005) and inflammatory bowel disease later in life (Le Huërou-Luron *et al.*, 2010).

[51] See for instance Patel *et al.* (2004) or Frech and Kimbro (2011).

[52] The pre-reform scheme of 12 weeks of unpaid leave was changed to 4 months of paid and 12 months of unpaid leave.

[53] Gender differences in parent-child bonding may rationalize the greater effects on MBDs for males. Murphy *et al.* (2010) document that daughters report higher levels of maternal affection and lower maternal overprotection, which lead to a lower incidence of MBDs in adulthood.

for men. Women, in contrast, are more likely to suffer from internalizing disorders, such as depression and anxiety, and thus may be less likely to end up being hospitalized (Federal Statistical Office, 2012). Furthermore, female prevalence rates of internalizing problems increase, but only at ages beyond the scope of this study.

Third, the increase in household income is another potential pathway for how the reform might affect child health outcomes. Whether ML expansions impact family income depends on the concrete context. In schemes with complete income replacements and no crowding out of unpaid leave, there is no effect on household resources when extending paid leave (Carneiro *et al.*, 2015, Dahl *et al.*, 2016, Bütikofer *et al.*, forthcoming). In contrast, a decrease in the amount of unpaid leave increases women's total earnings. This scenario is comparable with the ML expansion covered in this study. The 1979 ML reform raised mothers' average cumulative total income by 1,700 DM, with larger increases for women at the bottom of the wage distribution (see section 1.2). A considerable amount of research examines the positive impact of family income on children's cognitive achievement and health outcomes.[54] The underlying idea is that parents may invest more in their children because of the budget constraint's relaxation. Dahl and Lochner (2012) leverage changes in the earned income tax credit (EITC) and find that an increase in family income raises math and reading test scores, benefiting children from low socioeconomic statuses the most. Likewise, Hoynes *et al.* (2015) exploit expansions in the EITC and document that an increase in household income is accompanied by a reduced likelihood of low birth weight, which is induced by more prenatal care and less adverse health behavior such as smoking. Milligan and Stabile (2011) use variation in child tax benefits in Canada to investigate their impact on children's health outcomes in the Canadian context. They show quasi-experimental evidence that child benefits improve physical health for boys and mental health scores for girls. Aizer *et al.* (2016) study the long-run impact of the Mothers' Pension Program (1911-1935), a US state-funded welfare program with cash transfers to low-income families. They find that receiving cash transfers increased men's longevity. Akee *et al.* (2018) examine the impact of unconditional cash transfers on children. Drawing on the longitudinal dataset of the Great Smokey Mountains Study, the authors show that extra household income leads to a decrease in behavioral and emotional disorders, while, at the same time, childhood personality traits are improved. The children with the worst outcomes experience the largest gains, supporting the hypothesis that parents have a preference to equalize outcomes. As a potential mechanism, the authors report improved relationships between parents and children.

Since I find positive health differentials for children born under the more generous leave scheme, any of the previously discussed channels could be at play, as the reform improved children's environment in these contexts. However, because of the study's long-term nature and the lack of data on these mediating outcomes, this study cannot assess which mechanism is responsible for the downstream effects I observe.

[54] Recent literature has also been focusing on the associations between socioeconomic status and functional brain development (Tomalski *et al.*, 2013). In order to evaluate the causal effect of economic resources in early childhood on cognitive, socio-emotional, and brain development, the experiment *'Household Income and Child Development in the First Three Years of Life'* has been established by Greg Duncan and is running until 2022.

1.7 Concluding Remarks

In this paper, I analyze the impact of a ML reform on children's long-run health outcomes. To estimate causal effects for the length of ML, I use exogenous variation stemming from a legislative change in the Federal Republic of Germany in 1979, in which the length of paid ML was increased from eight weeks to six months after childbirth. I use hospital registry data over the period 1995-2014 in order to present a comprehensive analysis of the reform effect on important health outcomes. By following treated and untreated birth cohorts over 20 years of their adulthood (from age 16 up to age 35), I find evidence that the expansion in ML improves children's health in the long-run.[55] Children who were born after the implementation of the reform are on average 1.7 percent less likely to be hospitalized. There is strong heterogeneity by gender and age. The effects are mainly driven by men and the health differentials get stronger for those in their late 20s and after. Moreover, when looking into the components of hospitalizations, I observe that the decline in hospital admissions is due to fewer diagnoses of MBDs, the most common diagnosis type for individuals aged 15-35. Lastly, the largest effect of the ML reform on MBDs is observed for disorders due to psychoactive substance use and schizophrenia.

An interesting point for future research is to assess whether the effects also persist for more common diseases. The effects in the hospital registry data are likely just the *'tip of the iceberg'*, as they resemble rather extreme health outcomes. For this reason, it may be worthwhile to exploit health insurance data and investigate effects on health outcomes recorded by general practitioners.

From a policy perspective, this study suggests that a comprehensive cost-benefit analysis of leave schemes requires an understanding of the implications of ML on various aspects of child development. The primary reason for policymakers to extend the length of ML was to safeguard maternal well-being after childbirth. The results from this analysis show that the 1979 ML reform additionally led to positive health effects on children, but was only realized three decades later. This is an important takeaway for the assessment of such schemes. First, if the time horizon for assessing impacts is too short, certain consequences that develop only in the long-run may not be detected. Second, some benefits may not only materialize until later, but could also manifest in different areas than what policymakers had in mind initially. Therefore, if the vast effects on children's long-run health outcomes are not accounted for, cost-benefit analyses could come to wrong conclusions.

[55] As discussed in the paper, one caveat of the analysis is that the data only allows the reporting of ITT effects. The infeasibility to exclude unaffected individuals (e.g. foreign-born or born in the GDR) does not compromise the results but merely attenuates the ITT estimates. This dilution effect implies that the impact on treated children is greater than the reported ITT effects.

A Appendix to Chapter 1

A.1 Potential Threats to Identification and Validity of the Design

In the following, I discuss two threats to the empirical strategy. First, I examine potential self-selection into the treatment group by postponing delivery or strategic conception. Second, I consider threats along the distribution of birth months, which may result from seasonal and age of school entry effects.

Threats at the Threshold

Behavioral responses with respect to the forcing variable, the birth date of the child, could potentially jeopardize the validity of the identification strategy. Typically, gestation length is a normally distributed random variable with mean of 40 weeks and 2 weeks standard deviation, from the beginning of the last menstrual cycle (Ekberg *et al.*, 2013). However, parents may influence the time when their child is born in two ways.

First, parents may be incentivized to get strategically pregnant in response to the more attractive ML leave scheme. However, it is unlikely that children in my sample were conceived due to this reason. The draft bill was only proposed four months before the threshold birth date and the widest specification includes children who were born on October 1979, the month in which the earliest possible birth date could be. However, it is plausible that the topic increased public awareness due to media coverage even before the draft bill was initiated. For that reason, Dustmann and Schönberg (2012) conduct a literature search for articles regarding the reform. Subsequently, their results show that the earliest articles were published two months prior to the cutoff date.

Second, mothers with due dates close to the cutoff date on May 1, 1979 could have timed their child's date of birth by postponing induced births and cesarean sections.[56] Gans and Leigh (2009) find that during the introduction of a $3.000 'Baby Bonus' in Australia, parents postponed the birth by as much as a week in order to be eligible for the benefits: The data shows a sharp decline in the birth rate before the threshold, followed by a huge increase on the first day after the cutoff.

It is possible that there are similar distortionary 'introduction effects' in the context of the 1979 ML expansion. In other words, even though the announcement period does not allow for a strategic conception, parents may have been incentivized to delay the delivery of their child due to the reform.[57]

To check for such potential behavioral responses, I investigate the number of births around the cutoff. I follow the example of Gans and Leigh (2009) closely to show that there are no peculiarities in the fertility distribution. In the analysis, I use daily births from the federal states of Baden-Württemberg and North Rhine-Westphalia over the window of 1977-1990.

[56] Shifting a planned birth before the threshold is unlikely because this behavior potentially destroys mothers' eligibility for a more generous leave scheme.

[57] The scope of this effect is much smaller because the C-section or induced birth rates in the FRG around 1980 were significantly lower than in Australia in 2004. To be precise, Gans and Leigh (2009) report a C-section rate of 30% in Australia in 2004 while Selbmann and Thieme (1988) document a frequency of C-sections of 12.4% in Bavaria in 1982. As national numbers are not available for that time period, I use the second-largest state as a proxy for the federal level.

These two states are representative of the former FRG, as they accounted for almost 36% of all births in 1979. In the analysis, I limit the sample to a time window of one month before and after the cutoff date. Panel A of Figure A.1 shows the (unadjusted) daily number of births for April and May 1979. Over the time window, a strong weekly pattern with more births on weekdays and fewer births on weekends is observed. On May 1, which fell on a Tuesday, there was an unexpected drop in the number of births. If parents delayed the time of birth, one would expect a rise in the number of births right after the cutoff date. However, May 1 was a national public holiday (Labor Day), during which birth rates have been documented to be lower (Neugart and Ohlsson, 2013).

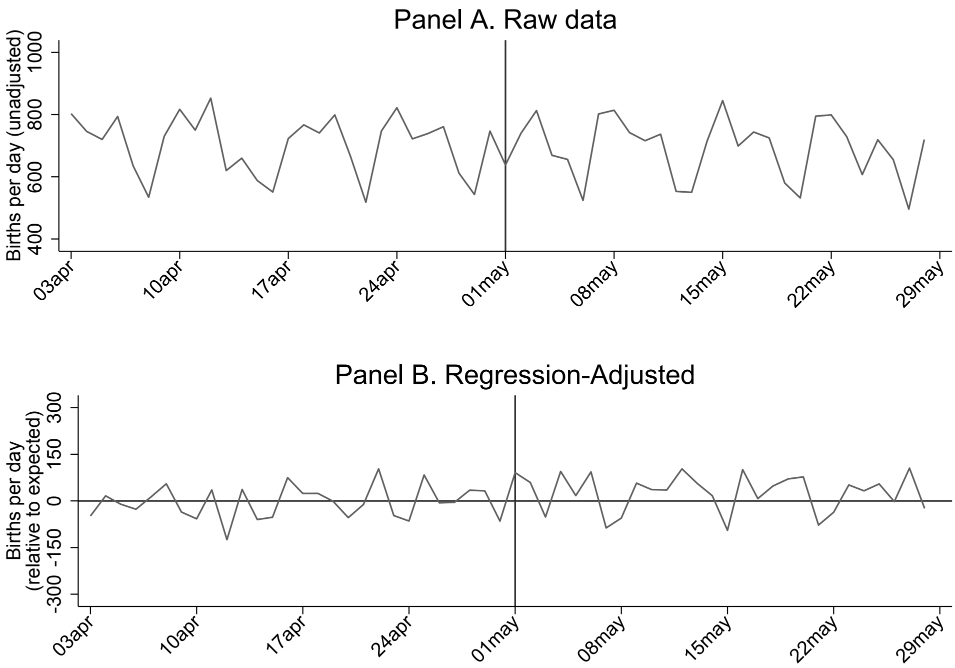


Figure A.1. Daily number of births around the ML expansion
Notes: The figure plots the number of births around the cutoff date May 01, 1979 for the ML expansion from two to six months after childbirth. Panel A shows the raw data, i.e. the actual number of births per day (unadjusted). Panel B, however, plots the difference between the raw and expected number of births when accounting for day of year, public holiday, and year×day of week fixed effects. For the expected number of births I use data in the same time window (one month before and after the threshold) for the years 1977-1990, except for the year in which the reform took place.
Source: Birth registry data from North Rhine-Westphalia and Baden-Württemberg. Taken together, both states account for almost 36% of all births in the former Federal Republic of Germany in 1979.

Panel B of Figure A.1 shows the time series after removing any variation in the timing of births stemming from year, day of week, day of year, and public holidays. To do so, I estimate the following equation on all years, *except* April and May 1979:

$$\text{Births}_i = I_i^{\text{Year}} \times I_i^{\text{Day of Week}} + I_i^{\text{Day of Year}} + I_i^{\text{Public Holiday}} + \varepsilon_i \qquad (1.3)$$

I regress the number of children born on day i, Births_i, on a series of dummies: year interacted with day of the week, day of the year, and a dummy for public national holidays.[58] Parameters are left out for better readability. Panel B plots the residuals ($\text{Births} - \widehat{\text{Births}}$)

[58]For public holidays I use Good Friday, Holy Saturday, Easter Sunday, Easter Monday, Labor Day, Ascension Day, Whit Sunday, Whit Monday, and Corpus Christi.

from the calibrated model. In contrast to Gans and Leigh (2009), I do not observe systematically fewer births prior to the reform, or more births right after the policy change came into effect. These results suggest that there is no evidence that parents delayed births to after the cutoff date.

Table A.1 presents estimates of the effect of the ML expansion on fertility outcomes, for different estimation windows.[59] Irrespective of whether the number of births or the log of the number of births is used, there is no evidence suggesting that parents deliberately postponed births until after the threshold date. On the contrary, the point estimates are negative throughout all specifications and become significantly different from zero with increasing estimation windows. Furthermore, the magnitude of the estimates is constant across estimation windows, which is a counterintuitive result if parents shifted births from the week prior to the reform to the week after the reform. If such behavior were present, one would expect point estimates to decrease in absolute value as the estimation window is enlarged.

Threats Along the Whole Distribution of Birth Months

As discussed in section 1.4, I cannot directly compare children born before and after the 1979 ML reform cutoff as birth months may affect outcomes directly (Buckles and Hungerman, 2013, Currie and Schwandt, 2013). Age-based cutoff rules at school entry are an important example in this respect (Black *et al.*, 2011). As the data is only available on the birth month level, I exploit a DiD design with different estimation windows around the cutoff, with a maximum bandwidth of six months on both sides of the threshold. Under the assumption that the treatment and control cohort share the same seasonality effects, the DiD design can isolate the reform impact from seasonality and age of school entry effects. There is no formal test to investigate whether the assumption of time-invariance of seasonality is met. However, the parallel trends for pre-threshold-born children in treatment and control group (Figures 1.2 and 1.4) provide evidence in favor of the assumption, suggesting that the cohort born in the year prior to the reform may be used as counterfactual.

The parallel trends cannot indicate whether post-threshold season of birth effects, such as discontinuities caused by school entry rules, may jeopardize the study design's validity. In West Germany, the cutoff date for school entry was June 30 and, consequently, falls into the post-threshold group of the 1979 ML reform (Jürges and Schneider, 2011).[60] If there were time-varying unobserved differences between children born around the school entry cutoff, the resulting DiD estimates would be biased. This may happen when treated children are affected differently by the school-based entry rules than control children. To check whether school entry rules or other forms of season of birth effects pose a threat to the identification, I run two robustness checks.

Exclude School Entry Discontinuity.—First, I limit the post-threshold group to children born before July (i.e. May and June). This way, school entry discontinuities are not part of the estimation. Figure A.2 shows estimates with varying bandwidths for the pre-cutoff side by gender and age group. The first estimate per panel, shown in green, rep-

[59] I estimate: $f(\text{Births}_i) = I_i^{\text{Reform}} + I_i^{\text{Year}} \times I_i^{\text{Day of Week}} + I_i^{\text{Day of Year}} + I_i^{\text{Public Holiday}} + \varepsilon_i$, with the dummy I_i^{Reform}, which equals one for days after the cutoff in 1979. The results remain qualitatively the same if $I_i^{\text{Public Holiday}}$ is interacted with $I_i^{\text{Day of Week}}$.

[60] The school entry rule dictates that children who turn six before to cutoff date, June 30, are admitted to primary school in that year. Children born after the threshold are admitted one year later.

Table A.1. Birth rate effects of the 1979 ML reform

	Estimation window			
	(1)	(2)	(3)	(4)
	±7 days	±14 days	±21 days	±28 days
Panel A. Dependent variable is number of births				
ML reform	-30.46	-30.23*	-33.32**	-32.78***
	(30.31)	(17.73)	(14.08)	(12.37)
Observations	196	392	588	784
R^2	0.856	0.842	0.832	0.817
Panel B. Dependent variable is ln(number of births)				
ML reform	-0.0448	-0.0440*	-0.0477**	-0.0476***
	(0.0425)	(0.0247)	(0.0197)	(0.0173)
Observations	196	392	588	784
R^2	0.855	0.844	0.833	0.819

Notes: The table shows regression estimates of the impact of the 1979 ML reform on fertility. The sample comprises daily births within the relevant estimation window in the federal states of Baden-Württemberg and North Rhine-Westphalia from 1977-1990. Panel A uses the number of births as dependent variable, whereas panel B shows results with the log of the number of births as dependent variable. All specifications control for day of year, public holiday, and year×day of week fixed effects. The estimation window is referring to the number of days before and after May 01. For instance, the ± 7 day window includes the last week in April and the first week in May across all years. Standard errors are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Birth registry data from North Rhine-Westphalia and Baden-Württemberg. Taken together, both states account for almost 36% of all births in the former Federal Republic of Germany in 1979.

resents the baseline estimate as shown in Tables 1.1 and 1.2. The pre-threshold estimation window contains half a year before the cutoff (Nov-Apr) in the widest specification and shrinks to two months before the threshold (Mar-Apr). The overall conclusions from the results section remain unaffected. The largest effects are found for the oldest age bracket (32-35 years) and males. The DiD estimates are of similar magnitude and mostly significant across the estimation windows. The only difference to the main results is that the effects for the age bracket of 27-31 years disappear.

Additional Control Cohort.—Second, I show estimates when using the additional control group born two years before the reform (i.e. born around May 1977, see column 5 of Table 1.6) for different bandwidths. If systematic month-of-birth patterns (including school entry discontinuities) varied over the cohorts, the estimates from this robustness check should differ from the baseline results. Figure A.3 presents estimates from this robustness check for various bandwidths. Once again, the estimates are of similar magnitude and significance compared to the baseline results. The reductions in hospitalizations are more pronounced in the older age brackets and for men. Related to this robustness check, it is worthwhile to point to the discussion on a potential violation of the *stable unit treatment value assumption* due to spillover effects on older siblings (see section 1.5.4). To refute this concern, I present estimates when using control cohorts from different birth years. The fact that the estimates have roughly the same magnitude irrespective of the control group does not support the idea that month-of-birth patterns are different through the years.

Overall, the additional robustness tests demonstrate that the main results are driven by the 1979 ML discontinuity and not by season of birth effects, such as school entry effects.

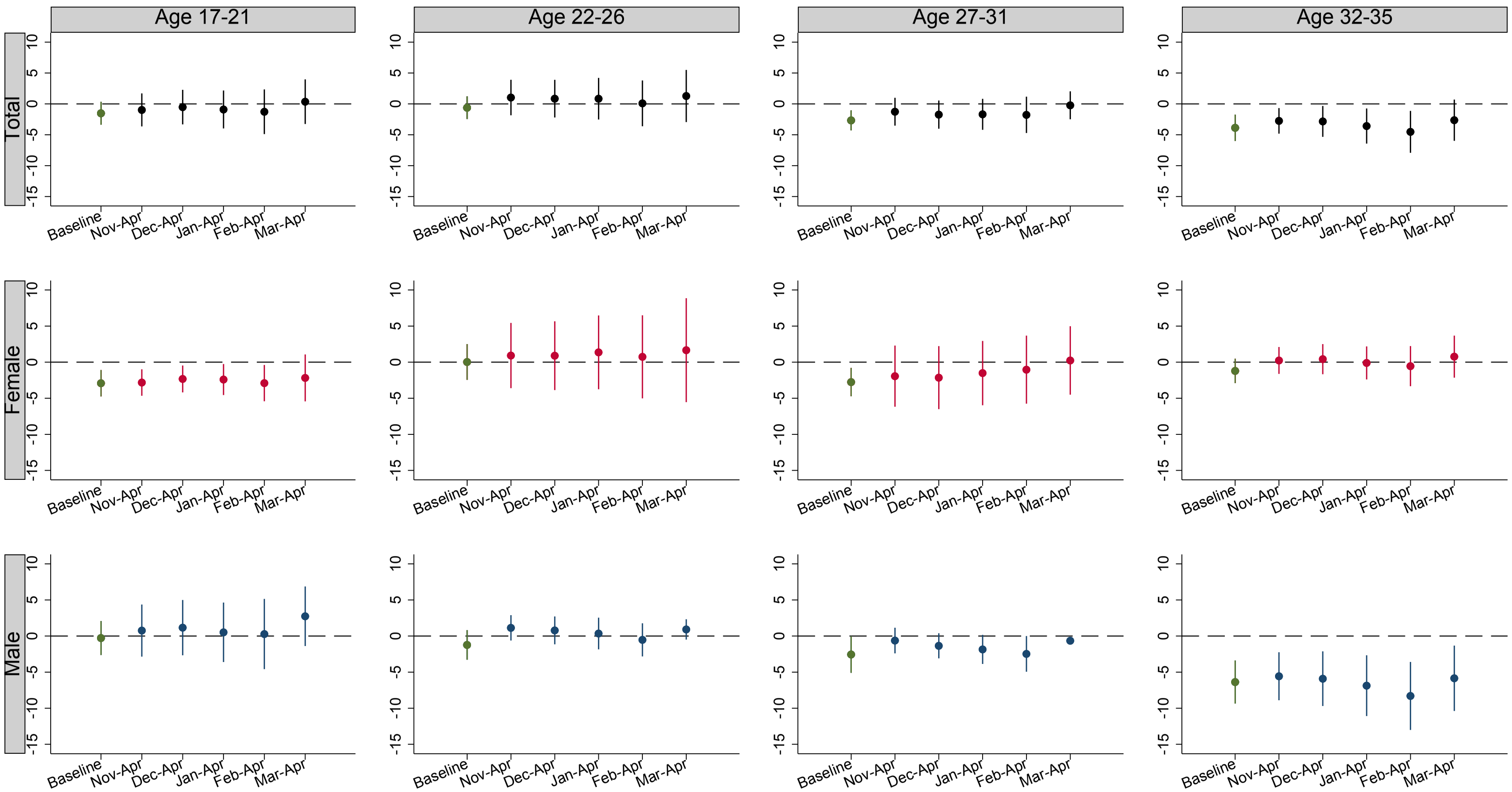


Figure A.2. Robustness—Account for school cutoff rules, by gender and age-group

Notes: The figure displays DiD estimates (along with 95% confidence intervals) of the 1979 ML reform on hospitalization. To check the robustness of the findings when accounting for school-entry rules, the post-threshold group is restricted to only include the months May and June. Each panel shows the effect of the reform when reducing the estimation window for the people born before the threshold (shrinking from half a year to two months). The first estimate per panel indicates the baseline estimate as shown in Table 1.1 and 1.2. It uses a bandwidth of half a year.

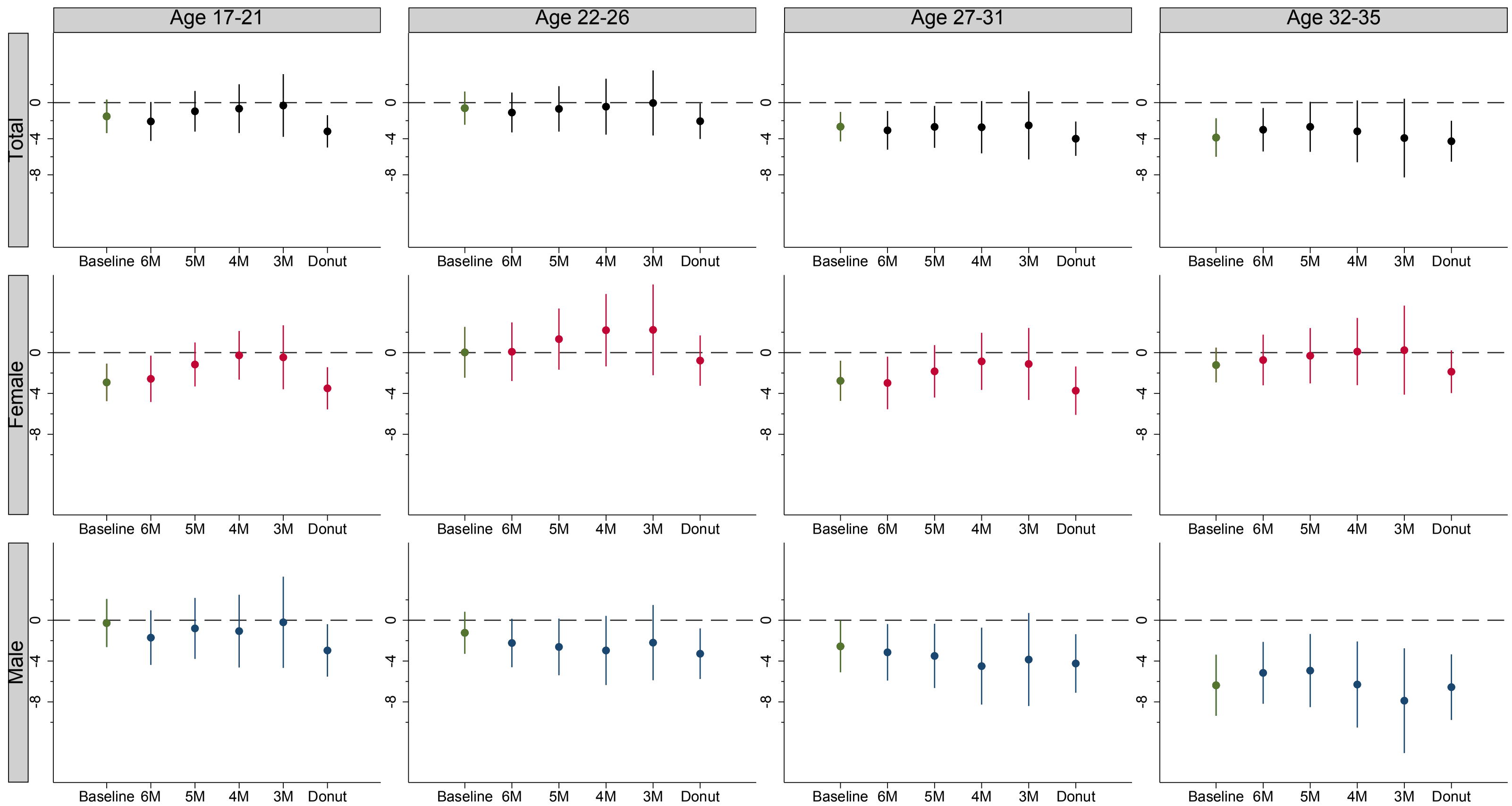


Figure A.3. Robustness—Using the additional control group, by gender and age-group

Notes: The Table shows DiD estimates (along with 95% confidence intervals) of the 1979 ML reform on hospital admissions (pooled). To check the robustness of the findings when accounting for school-entry rules, it presents estimates when including the additional control group born two years before the reform (i.e. born around May 1977, see column 5 of Table 1.6). The DiD estimates are reported for different estimation windows around the cutoff. The *'Donut'* specification uses a bandwidth of half a year and excludes children born in April and May. The first estimate per panel indicates the baseline estimate as shown in Table 1.1 and 1.2. It uses a bandwidth of half a year.

A.2 Figures

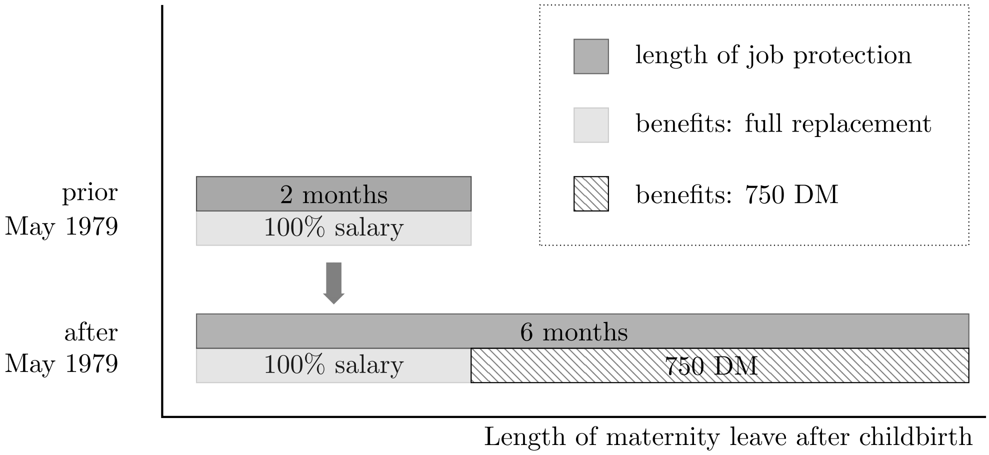


Figure A.4. 1979 reform in ML legislation in the Federal Republic of Germany
Notes: The figure describes the legislative change in the length of job protection and ML, which took place in the Federal Republic of Germany in 1979. The reform increased post-birth ML from eight weeks to six months, while keeping the initial structure of the period from six weeks before until eight weeks after childbirth unchanged (mother protection period).
Source: The figure is based on information from Dustmann and Schönberg (2012), Ondrich *et al.* (2002), Schönberg and Ludsteck (2014) as well as Zmarzlik *et al.* (1999).

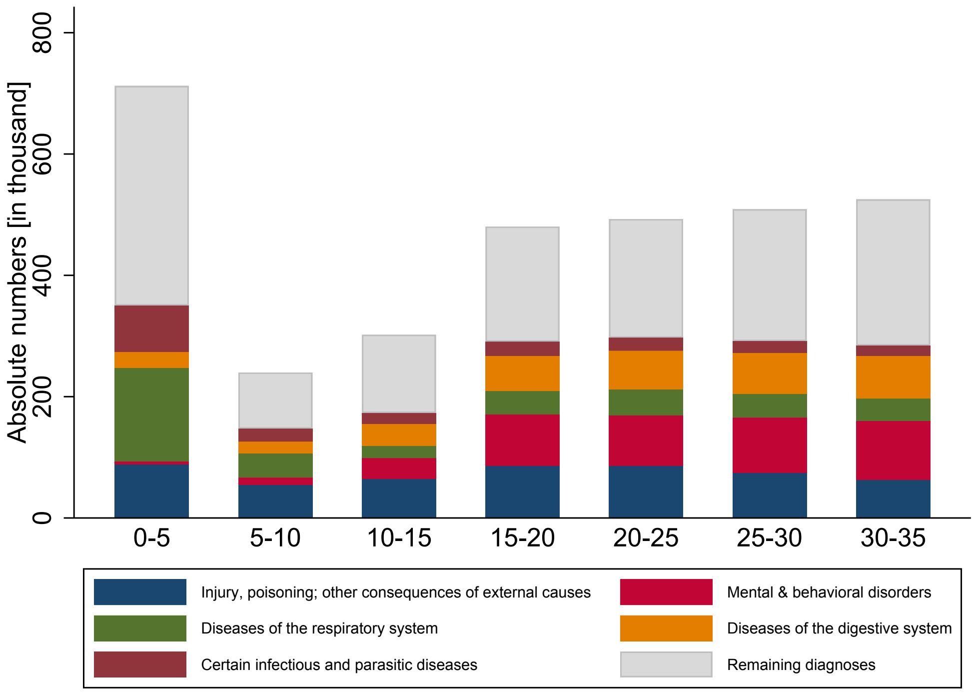


Figure A.5. Five main diagnoses of inpatients aged 0 to 35 in 2014
Notes: The figure shows the incidence distribution across different age brackets of the top five diagnoses for inpatients aged 0 to 35 in 2014. Diagnoses associated with pregnancy, childbirth, and the puerperium are not taken into account in this representation. The large remainder in the age bracket '0-5' consists mostly of conditions that originate in the perinatal period.

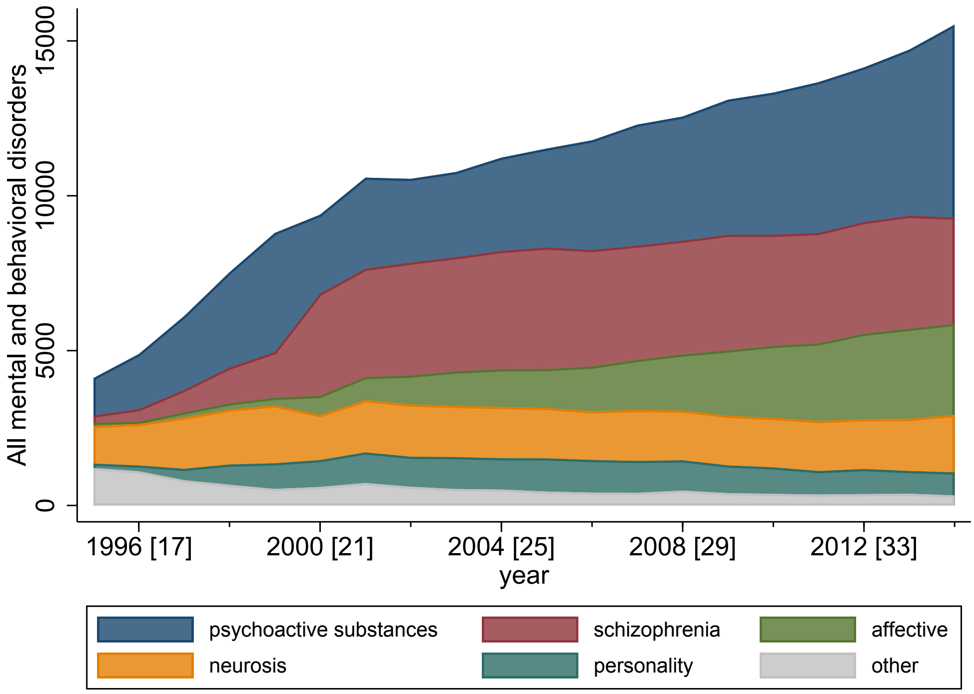


Figure A.6. The top five subcategories of mental and behavioral diagnoses
Notes: This figure plots the yearly number of diagnoses for the treatment cohort (i.e. the individuals born between November 1978 and October 1979). The subcategories are ordered by their occurrence in 2014 (from the most to the least frequent diagnosis), which also coincides by chance with the ordering in the ICD-10 classification system. The five most frequent subcategories - as shown here - comprise more than 95% of all MBDs.

(a) Region boundaries **(b)** Population density

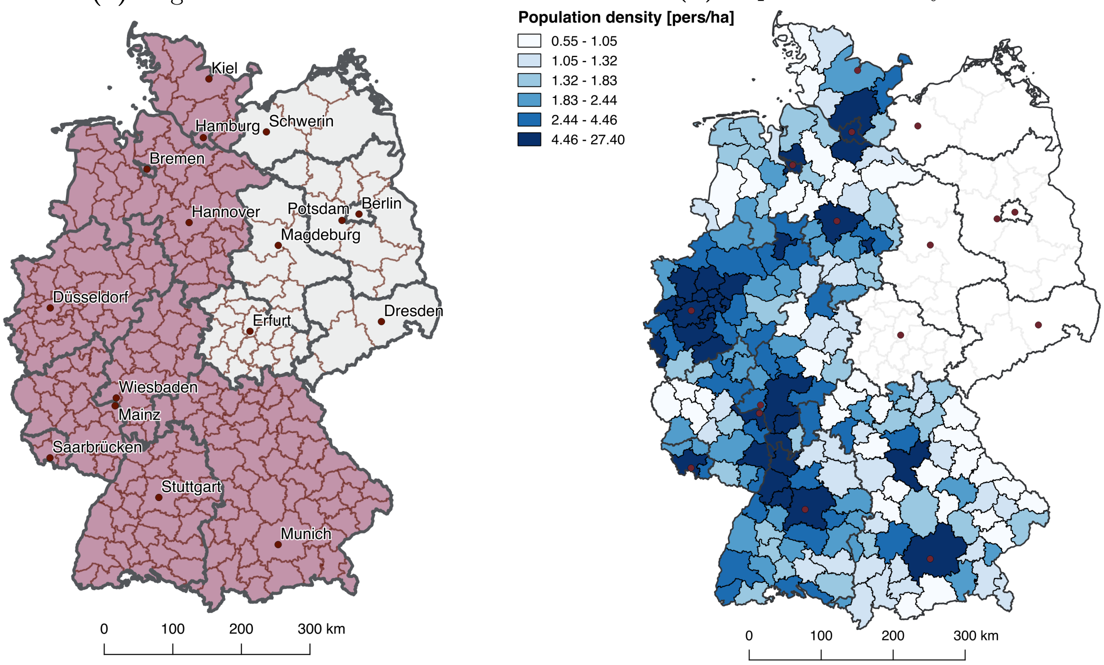


Figure A.7. Labor market regions in Germany

Notes: The map on the left shows the labor market regions (LMR) used in the analysis. The areas with the red background depict the area of the former Federal Republic of Germany ('West Germany'), while the white areas indicate the area of the former German Democratic Republic ('East Germany'). The area of West Germany is used throughout the paper, the regions of East Germany only in a robustness check (triple-differences model). The baseline specification aggregates to level of West and East Germany, yet there are some specifications that aggregate to the regional level (red borderlines). There are in total 245 LMR, with 204 in the area of the FRG and 41 in the area of the former GDR. The black outlines indicate federal state boundaries and the red dots represent the corresponding state capitals. The map on the right presents the regional variation of population density across German regions. Labor market regions are labeled as urban if their population density exceeds the median value of all regions.

Source: Own representation with data from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).

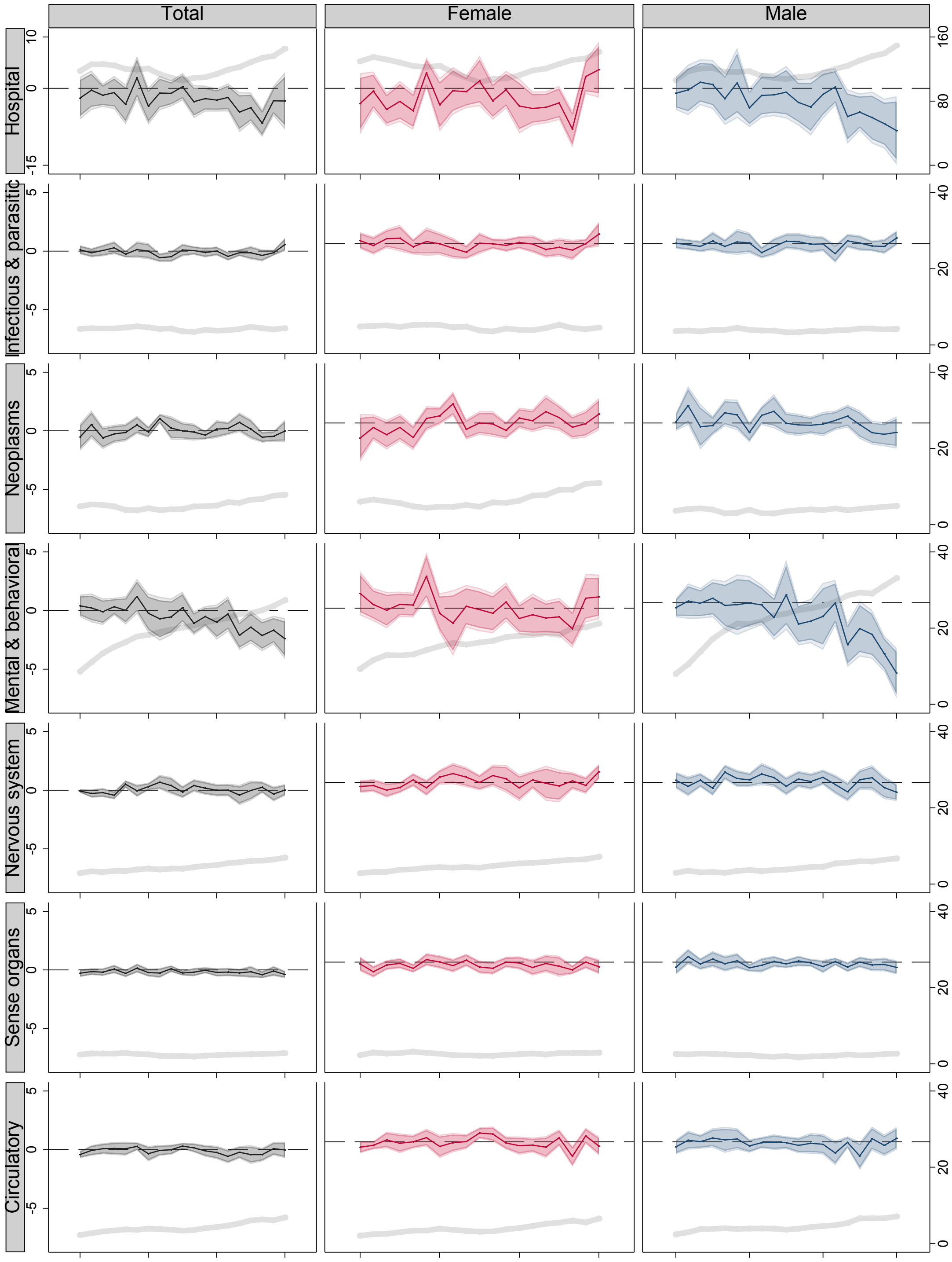
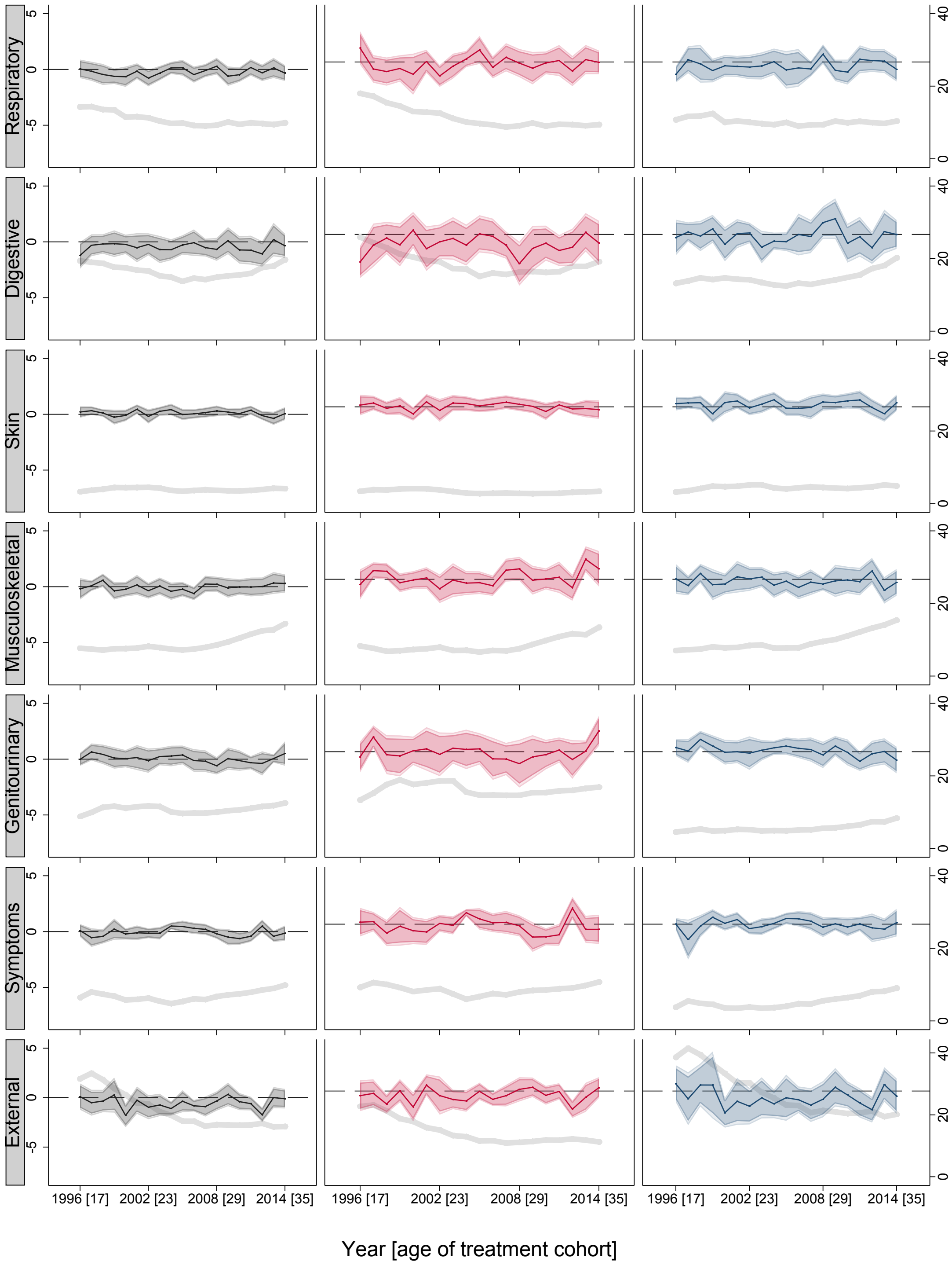


Figure A.8. Life-course approach for all chapters
Continued on next page

Life-course approach for all chapters (continued)



Notes: This figure plots intention-to-treat estimates (along with confidence intervals) across the main diagnosis chapters for the entire life-course. The outcomes are defined as the number of cases per 1,000 individuals (births). The point estimates are coming from a DiD regression as described in section 1.4, with a bandwidth of six months, month-of-birth fixed effects, and clustered standard errors on the month-of-birth level. The control group is comprised of children that are born in the same months but one year before (i.e. children born between November 1977 and October 1978). On the right axis, one can see the dependent mean for the pre-reform treatment children.

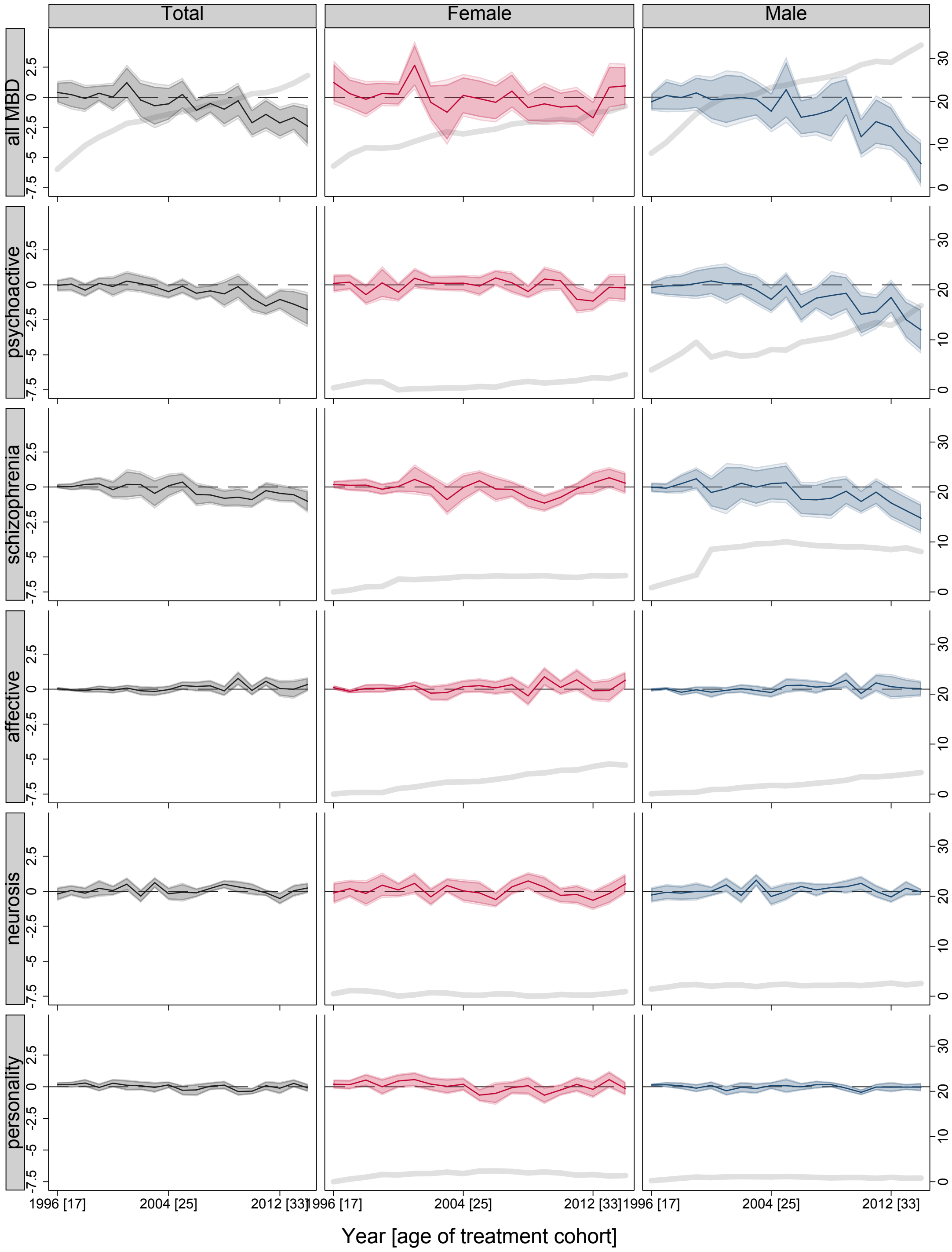


Figure A.9. Life-course approach for the subcategories of mental and behavioral disorders.
Notes: This figure plots intention-to-treat estimates (along with confidence intervals) across the subcategories of MBDs. The outcomes are defined as the number of cases per 1,000 individuals (births). The point estimates are coming from a DiD regression as described in section 1.4, with a bandwidth of six months, month-of-birth fixed effects, and clustered standard errors on the month-of-birth level. The control group is comprised of children that are born in the same months but one year before (i.e. children born between November 1977 and October 1978).

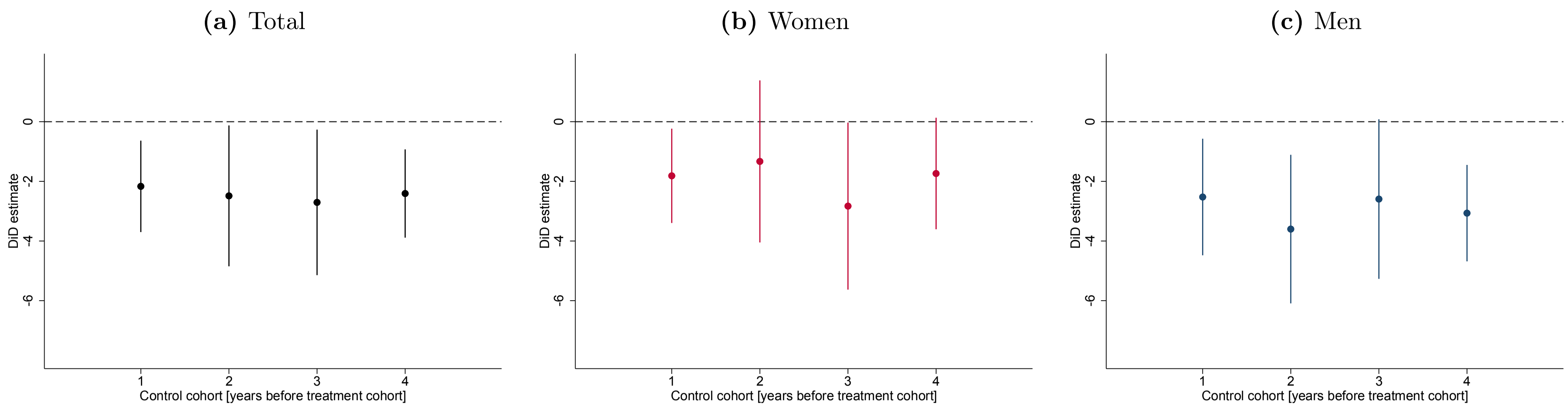


Figure A.10. Potential spillover effects on older siblings for hospital admission
Notes: The figure shows DiD estimates for the effect of the 1979 ML reform on hospital admissions when using different control cohorts. Each estimate corresponds to a different regression when exchanging the control groups and not accounting for age differences. For instance, the control cohort two years before the treatment cohort contains individuals born between November 1976 and October 1977. All regressions use a bandwidth of half a year. The outcomes are defined as the number of hospitalizations per 1,000 individuals. Column a displays the results for all admissions, whereas columns b and c show the estimates for women and men, respectively.

A.3 Tables

Table A.2. Summary statistics for different diagnoses

	(1) ICD-9	(2) ICD-10	(3) Mean	(4) SD
A. Hospital admission			120.625	10.961
Infectious and parasitic diseases	001-139	A00-B99	4.210	0.493
Neoplasms	140-239	C00-D48	5.155	1.282
Mental & behavioral disorders	290-319	F00-F99	18.956	5.548
Diseases of the nervous system	320-359	G00-G99	4.500	1.264
Diseases of the sense organs	360-389	H00-H95	2.404	0.348
Diseases of the circulatory system	390-459	I00-I99	4.108	1.380
Diseases of the respiratory system	460-519	J00-J99	10.994	1.939
Diseases of the digestive system	520-579	K00-K93	16.746	2.079
Diseases of the skin and subcutaneous tissue	680-709	L00-L99	3.849	0.536
Diseases of the musculoskeletal system	710-739	M00-M99	8.897	2.228
Diseases of the genitourinary system	580-629	N00-N99	10.621	1.362
Symptoms, signs, and ill-defined conditions	780-799	R00-R99	6.794	1.410
Injury, poisoning and certain other consequences of external causes	800-999	S00-T98	21.196	5.978

Continued on next page

Summary statistics for different diagnoses (continued)

	(1) ICD-9	(2) ICD-10	(3) Mean	(4) SD
B. Mental & behavioral disorders			18.956	5.548
Organic, including symptomatic, mental disorders	290,293,294,310	F00-F09	0.115	0.056
MBD due to psychoactive substance use[1]	291,292,303,304,305	F10-F19	6.366	2.232
Schizophrenia, schizotypal and delusional disorders	295,297,298	F20-F29	5.140	2.246
Mood [affective] disorders	296,311	F30-F39	2.339	1.673
Neurotic, stress-related and somatoform disorders	300,306,308,309	F40-F48	2.799	0.356
Behavioural syndromes associated with physiological disturbances and physical factors	316	F50-F59	0.308	0.225
Disorders of adult personality and behavior	301,302	F60-F69	1.375	0.511
Mental retardation	317,318,319	F70-F79	0.121	0.075
Disorders of psychological development	299,315	F80-F89	0.026	0.029
Behavioural and emotional disorders with onset usually occurring in childhood and adolescence	312,313,314,307	F90-F98	0.320	0.535

Notes: The table shows the classification of diseases according to the 'International Statistical Classification of Diseases and Related Health Problems (ICD)', a medical classification list provided by the World Health Organization. For the main chapters, the mapping between ICD-9 and ICD-10 is taken from the European shortlist. The table reports next to the ICD codes summary statistics for the different diagnosis types. Column 3 and 4 show mean and standard deviation of the number of diagnoses per 1,000 individuals for the pre-reform treatment cohort. The data uses the modified version of the ICD system issued by DIMDI.
Source: The ICD coding is taken from the World Health Organization (WHO), see for example: http://www.who.int/classifications/icd/en/ and the European shortlist (Federal Statistical Office, 2012, p. 76), the summary statistics are obtained from the hospital registry data.
[1] Psychoactive substances include alcohol, opioids, cannabinoids, sedatives or hypnotics, cocaine, other stimulants (including caffeine), hallucinogens, tobacco, volatile solvents, multiple drug use and use of other psychoactive substances.

Table A.3. RDD on hospital admissions

	Estimation window		
	(1) 6M	(2) 5M	(3) 4M
Panel A. Over entire length of the life-course			
Overall	-2.818	-2.148	-2.084
	(4.230)	(4.976)	(5.848)
Panel B. Age brackets			
Age 17-21	-2.870	-2.509	-3.697
	(4.181)	(5.162)	(6.081)
Age 22-26	-1.844	-1.084	-2.049
	(4.254)	(5.151)	(6.224)
Age 27-31	-2.083	-1.450	-0.434
	(4.312)	(4.919)	(5.751)
Age 32-35	-4.889	-3.900	-2.173
	(5.030)	(5.691)	(6.495)

Notes: The Table shows RDD estimates coming from the following specification: $Y_{mt} = \beta_0 + \beta_1 After_m + \beta_2 \tilde{X}_m + \beta_3 After_m \times \tilde{X}_m + \varepsilon_{mt}$. $\tilde{X}_{im} = (X_{im} - c)$ *corresponds to the normalized birth month and* β_1 *is the parameter of interest. The sample includes the treatment group (individuals born around the threshold) only. Standard errors clustered on the month-of-birth level.*

Table A.4. Number of observations per age-bracket

	Estimation window				
	(1) 6M	(2) 5M	(3) 4M	(4) 3M	(5) Donut
Age 17-21	120	100	80	60	100
Age 22-26	120	100	80	60	100
Age 27-31	120	100	80	60	100
Age 32-35	96	80	64	48	80

Notes: The table reports the number of observations per age bracket and estimation window. It refers to the DiD estimations shown in Panels B of Tables 1.1, 1.2, 1.4, and 1.5.

Table A.5. Robustness—Interaction treatment with age brackets (hospitalization)

	Estimation window				
	(1) 6M	(2) 5M	(3) 4M	(4) 3M	(5) Donut
Panel A. Total					
Treatment × Age 17-21	-2.310***	-2.143**	-2.156*	-2.386*	-2.647***
	(0.729)	(0.829)	(1.039)	(1.309)	(0.715)
Treatment × Age 22-26	-0.0735	0.0501	-0.0501	-0.00689	-0.451
	(0.764)	(0.916)	(1.149)	(1.422)	(0.788)
Treatment × Age 27-31	-2.187**	-1.789	-2.350	-2.323	-2.644**
	(0.977)	(1.134)	(1.379)	(1.696)	(1.064)
Treatment × Age 32-35	-4.148***	-4.040***	-4.639***	-4.620**	-5.060***
	(1.021)	(1.164)	(1.390)	(1.747)	(1.015)
Dependent mean	121.1	121.0	121.5	123.3	121.9
N (MOB × year)	456	380	304	228	380
Panel B. Women					
Treatment × Age 17-21	-0.879	-0.607	0.153	-0.292	-1.370
	(0.992)	(1.048)	(1.176)	(1.560)	(0.946)
Treatment × Age 22-26	0.206	0.880	1.487	1.234	0.177
	(0.943)	(1.058)	(1.278)	(1.698)	(0.872)
Treatment × Age 27-31	-3.083***	-2.204*	-1.923	-2.429	-3.404***
	(1.060)	(1.119)	(1.338)	(1.790)	(1.035)
Treatment × Age 32-35	-3.580***	-3.401**	-2.921*	-2.239	-4.533***
	(1.191)	(1.347)	(1.586)	(1.830)	(1.153)
Dependent mean	122.3	121.9	121.9	123.8	123.2
N (MOB × year)	456	380	304	228	380
Panel C. Men					
Treatment × Age 17-21	-3.738***	-3.635***	-4.360***	-4.397**	-3.937***
	(0.984)	(1.110)	(1.317)	(1.561)	(1.133)
Treatment × Age 22-26	-0.352	-0.751	-1.513	-1.193	-1.060
	(1.072)	(1.264)	(1.521)	(1.696)	(1.212)
Treatment × Age 27-31	-1.325	-1.402	-2.759	-2.224	-1.908
	(1.318)	(1.585)	(1.811)	(1.914)	(1.548)
Treatment × Age 32-35	-4.679***	-4.648***	-6.275***	-6.887***	-5.551***
	(1.311)	(1.545)	(1.666)	(2.063)	(1.440)
Dependent mean	120.0	120.2	121.2	122.7	120.7
N (MOB × year)	456	380	304	228	380

Notes: The table reports DiD estimates when using interactions of $Treat \times After$ with the age brackets and relying on the full sample. This robustness check differs from the specification in Table 1.1 in which there are different regressions for each age-bracket.

Table A.6. ITT effects on the subcategories of mental and behavioral disorders

	(1)	(2)	(3)	(4)	(5)
	Overall	Age brackets [years]			
		17-21	22-26	27-31	32-35
MBD	-0.621**	0.174	-0.008	-1.000***	-1.906***
	(0.242)	(0.257)	(0.410)	(0.349)	(0.362)
Psychoactive substances	-0.483***	-0.074	-0.071	-0.549***	-1.428***
	(0.110)	(0.123)	(0.136)	(0.156)	(0.270)
Schizophrenia	-0.272**	0.061	0.069	-0.707***	-0.572***
	(0.119)	(0.087)	(0.230)	(0.155)	(0.170)
Affective	0.093**	-0.035	0.004	0.198***	0.235*
	(0.035)	(0.042)	(0.054)	(0.068)	(0.128)
Neurosis	0.066	0.001	0.108	0.213***	-0.088
	(0.040)	(0.086)	(0.102)	(0.066)	(0.054)
Personality	0.013	0.172***	0.005	-0.158*	0.039
	(0.036)	(0.055)	(0.072)	(0.086)	(0.091)
N (MOB $\times$ year)	456	120	120	120	96

Notes: The table shows DiD estimates of the 1979 ML reform on subcategories of MBDs. The first column shows the effect for the entire pooled time frame, whereas columns 2 to 5 display the impact per age group. Each row corresponds to a different estimation with the number of diagnoses per 1,000 individuals. See Table 1.1 for additional details. Clustered standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7. Robustness checks for mental and behavioral disorders

		Alternative specifications		Alternative estimation		Placebos	
	(1) Baseline	(2) Current population	(3) LMR level[a]	(4) DDD[b]	(5) Add. CG	(6) Temporal: cohort	(7) Spatial: GDR
(1) Total	-0.621**	-0.832***	-0.844***	-0.872**	-0.553**	0.252	0.252
	(0.242)	(0.239)	(0.219)	(0.321)	(0.269)	(0.320)	(0.155)
(2) Female	0.010	-0.0853	-0.130	-0.0138	0.289	0.527	0.0235
	(0.271)	(0.266)	(0.261)	(0.326)	(0.287)	(0.392)	(0.215)
(3) Male	-1.192***	-1.554***	-1.558***	-1.627***	-1.323***	-0.005	0.434*
	(0.288)	(0.299)	(0.283)	(0.392)	(0.322)	(0.334)	(0.225)
For total:							
Dependent mean	19.57	17.28	17.88	19.57	18.96	19.21	8.850
Effect in SDs [%]	12.44	41.92	5.230	17.48	9.960	6.120	12.91
N	456	288	53,855	912	720	456	456
MOB fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: This table displays robustness checks for the effect of the 1979 maternity leave reform on mental and behavioral disorders. I perform the following checks (with reference to the column): (1) baseline specification that was used in previous parts of the paper, (2) for the outcome I use the number of diagnoses divided by the current number of individuals (approximation), (3) the analysis is carried out on the level of labor market regions, (4) triple difference model (the third difference stems from the former region of the GDR), (5) I use as control cohort not only the cohort before the reform, but also the cohort 2 years prior to the policy change, (6) first placebo, in which the entire analysis set-up is pushed back by one year, i.e. the placebo TG is the cohort prior to the real TG and the placebo CG is the cohort born 2 years before the reform took place, (7) second placebo, in which I run the normal DD set-up in the area of the former GDR.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[a]: level of analysis on Labor Market Regions: weighted regressions (by population), includes region fixed effects.

[b]: standard errors clustered on the month-of-birth×birth-cohort×East-West cell level.

Chapter 2

Crime Externalities from Football Games in Germany

2.1 Introduction

Football has grown from being a favorite sport in many parts of the world into a respected industry, which is reflected by high attendance rates and considerable revenues.[1] For instance, the *Bundesliga*, the German top league, is a major reason for public mass gatherings and had with 42,700 spectators the highest average attendance per match of all European leagues in the season 2017/18. Also in economic terms, football is synonymous with big business as the *Bundesliga* ranks second in Europe with generated revenues of 3.17 billion euros (Deloitte, 2019). Due to their great popularity, professional football matches generate external effects in many regards. On the positive side, there is higher consumer spending (e.g. merchandising, catering, and accommodation) and ultimately increased local tax revenues. On the negative side, the games induce air and noise pollution, and a higher frequency of accidents, among other things. Violent fan behavior is a particularly pervasive aspect in this context and affects individuals' health, the penal system, and the police. In the 2017/18 season, the State and Federal Police force invested over 2.11 million working hours to ensure the spectators' safety at professional football matches in Germany (Central Sports Intelligence Unit, 2018). There has been a widely noticed legal dispute between the Federal State of Bremen and the German Football League (DFL) regarding the reimbursement of costs for high-risk games, in which the Federal Administrative Court (2019) declared the charges imposed by Bremen as legitimate. This study does not only focus on football matches where social interactions lead to violence, but also deals with the question of whether the negative externalities of football games that bring so much prosperity should be absorbed.

The aim of this paper is to examine in detail the effects of football games on violent behavior in the German professional football league. To assess the causal impact of football matches on physical assaults, I employ a generalized difference-in-differences approach that exploits the variation in the timing of the football games. I compare the level of assaults on days with and without home games while controlling for any potential source of heterogeneity across days of the week, month, and year, and taking into account other

[1] Except where otherwise noted, football refers to the sport also known as soccer.

possible confounding variation from weather or holidays. In order to use the time-series and cross-sectional variation in the data, I construct a panel at the municipality-day level. I collect and match data from various sources to analyze the effect of 4,461 football games in the period 2011-2015. First, detailed information on all matches in the first three leagues is obtained via web scraping.[2] Second, the primary outcome variable, the rate of physical assaults, is derived from comprehensive registry data provided by the Federal Criminal Police Office. It covers all individuals that become victims of a crime against their legally protected rights in the sample regions. Third, data from weather monitors and a time-series of holidays at the state level are matched to the data to account for possible confounders. Last, population figures from the Federal Statistical Office are merged to construct regression weights and create the outcome variable.

I find significant and robust evidence that football matches lead to an increase in the prevalence of physical assaults. A home game increases the assault rate by on average 21.5 percent. The effects are almost entirely driven by male victims and are most pronounced in the 18-29 age group. Moreover, a heterogeneity analysis reveals that the effects are caused by attacks on victims with no prior relationship to the suspect and completed offenses (i.e. no attempts). Assaults on police officers on duty account for 16 percent of the baseline effect of the entire sample. I find no evidence that the effects of football games on the prevalence of assaults are offset by reductions in violent acts on days adjacent to the game day or in nearby areas. The focus on home games in the empirical framework does not pose a threat to the identification strategy, as away games do not change the level of physical assaults in the home district.

There are a number of theories rationalizing spectator violence. For instance, the frustration-aggression hypothesis, proposed by Dollard *et al.* (1939), interprets violent behavior as an act to rehabilitate self-esteem triggered by a frustrating event, such as the defeat of one's favorite team. Another prominent theory is the idea of social learning, according to which spectators mimic the behavior displayed on the field (Bandura, 1973, 2007). In my analysis, I test whether the frustration-aggression hypothesis can serve as an explanation for the increase in physical assaults. To do so, I use emotional cues from either emotionally unsettling events during the game or when game outcomes do not align with pregame expectations. In both cases, there is no evidence to support the frustration-aggression hypothesis. In the second step, I consider the prominence of games and teams as potential channels. In games with high rivalry, the impact on the rate of physical assaults is three times higher. This larger effect may be related to the mutual dislike of opposing fan groups or to the social learning theory, as high-rivalry matches are characterized by more intensity on the field. Furthermore, I find larger effects for the highest division of the league system, which attracts the most viewers. Ultimately, I cannot pinpoint the one channel at play, but there is suggestive evidence that spectators mimic the behavior of the players, that violent viewers select into specific matches, and eventually a mechanical agglomeration effect.

Back-of-the-envelope calculations based on my estimates indicate that the crime externalities of football games are substantial. In the season of 2014/15, for example, football games in the top three leagues of the German football league system explain 17.7 percent of assault reports and lead to 18,770 additional assaults in the sample regions. Glaubitz

[2]This includes, but is by no means limited to the time and location of the games. Furthermore, I exploit betting odds that reflect pregame expectations.

et al. (2016) estimate the social cost of an assault at 5,067 euros (in 2020 prices), which implies an annual social cost of 95 million euros.[3]

This paper complements the literature on sporting events and their effects on different dimensions of crime. Studies in the US American context demonstrate that American college football games lead to a higher incidence of rape (Lindo *et al.*, 2018), physical assaults, vandalism, disorderly conducts, and alcohol-related offenses (Rees and Schnepel, 2009). Furthermore, it has been documented that intimate partner violence is linked to unsettling defeats in professional American football games (Card and Dahl, 2011). In the European setting, studies have shown that professional football games increase property and violent crimes (Marie, 2016), and shift offenses both temporally (Montolio and Planells-Struse, 2016) as well as spatially (Montolio and Planells-Struse, 2019).

The empirical analysis in this paper contributes to previous literature in many regards. I compile a unique data set from various sources to comprehensively characterize the impact of football matches on criminal behavior in the top three German professional football divisions. Using the universe of criminal offenses, I show that almost one in five physical assaults can be attributed to professional football games. In addition to the overall effect, I show effect heterogeneity by victim and crime characteristics. When investigating channels, I find no support for the frustration-aggression hypothesis, but evidence for the social learning theory and a selection story in which violent fans self-select into specific matches.

The remainder of the paper is structured as follows. The next section provides information about football games, their relationship with violent spectator behavior, and previous literature. Section 2.3 explains the data and the variables. Section 2.4 contains a description of the empirical framework. Section 2.5 reports results, validity checks, a discussion on potential channels, and robustness tests. Section 2.6 concludes.

2.2 Background

2.2.1 The German Football League System

The three fully professional divisions in the German football league system are managed under the jurisdiction of the German Football Association (DFB) and the German Football League (DFL). While the top two leagues, *Bundesliga* and *2. Bundesliga*, are organized by the DFL, the third division, *3. Liga*, is run by the DFB itself. Teams can be promoted or relegated from one league to another. The top two divisions consist of 18 teams playing 17 home and away games in one season. The third league contains 20 teams playing 19 home and away games.

The empirical approach of this paper exploits the variation in the scheduling of matches. Since 2006, the match schedules for the *Bundesliga* and the *2. Bundesliga* are created with a software that uses integer linear programming.[4] The software outlines the rough details

[3]The estimated social costs of Glaubitz *et al.* (2016) are conservative compared to estimates from other countries. In the United States, the social costs of an assault range from 17,300 to 68,000 euros (Miller *et al.*, 1996, Cohen *et al.*, 2004). In New Zealand the social costs are estimated at 6,400 euros (Roper and Thompson, 2006), and in Great Britain the estimated costs are 2,300 euros (Dubourg *et al.*, 2005). All costs are in 2014 prices.

[4]For details on how the match schedules are created, please refer to `https://www.bundesliga.com`.

such as the matches per gameday. The exact date and time, however, are determined in the course of the season. The later exact scheduling makes it possible to take into account guidelines from local authorities, security bodies, the Central Sports Intelligence Unit (ZIS), international football associations (FIFA/UEFA), fans, clubs, and stadium operators. In addition to obvious restrictions such as the fact that home games of neighboring clubs should be scheduled at different times, the DFL has to consider public holidays, other major events, or match dates of international competitions.

2.2.2 Football and Violent Crime

Spectator violence has a long tradition in the context of professional football in Germany. The change in names for football fans illustrates that spectator behavior has changed considerably (Pilz, 2005). In the 1960s and 1970s, the peaceful fan base was refereed to as *camp-followers*, while one decade later the first problems of spectator violence emerged with the so-called *football rowdies*. In the 1980s, spectator violence was omnipresent, mainly due to the hooligan movement. Since the late 1990s, a new group has appeared in the stadiums, the *ultras*. Originally from Italy, the *ultras* are dedicated to fighting the commercialization of football and to revitalizing traditional football culture. Over the last years, the number of violent fans has been increasing. The police distinguish between three types of football fans. Category A includes peaceful fans, category B consists of fans inclined to violence, and category C contains fans who actively seek violence (violent criminals). Originally, the *ultras* were predominantly assigned to category A and occasionally to category B. Recently, however, a substantial share of the *ultras* has been classified as members of categories B and C. In the season 2017/18, for instance, the Central Sports Intelligence Unit (2018) identified 13.633 individuals in the top three leagues who were either prone to violence or seeking violence. With more potential agitators, police officers report more aggression directed at them (Feltes, 2010). However, it is not only the growing number of violent fans that is a problem, but also the fact that the event character of violence during football games is increasing (Pilz, 2005). Violent actions are less and less connected with the events on the playing field. In fact, violence serves as an opportunity for identification in the search for one's identity. Some violent fans find self-assertion by joining like-minded individuals to discover their strength. The question is how to address the challenges posed by *hooligans* and *ultras*. Poutvaara and Priks (2009) investigate the impact of anti-hooliganism policy in the Swedish context. They find that the assignment of other duties to the Sport Intelligence and Tactical Unit was accompanied by an increase in violent crime by hooligans. Feltes (2010) emphasizes the concept of 'balanced policing', which acknowledges that risks from spectators are dynamic and require quick responses to either escalate or deescalate.[5]

There are a number of theories rationalizing spectator violence.[6] First, the higher number of assaults could be a merely mechanical effect resulting from increased social contacts and a change in the way interactions take place (Montolio and Planells-Struse, 2019). For

[5]Depending on the types of violent fans the police are confronted with, the officers have to adjust their behavior accordingly (Feltes, 2010). On the one hand, a zero-tolerance policy must be applied to *hooligans* and any failure to take drastic measures is seen as an invitation to act out violent desires. On the other hand, *ultras* should be given some leeway to allow self-regulation.

[6]Theoretically, football games can also lead to a decrease in the assault rate. This may happen if individuals with a higher propensity for violent behavior attend the matches. Dahl and DellaVigna (2009) find a similar effect for violent movies and refer to this as the 'self-incapacitation effect'. However, this effect can only be demonstrated with hourly data. As my analysis draws on daily data, I focus on theories as to why the rate of violence increases due to football games.

instance, the *Bundesliga* has the highest average number of spectators per match of all European major football leagues (Wicker *et al.*, 2017). In the 2014/15 season, an average of more than 42,000 spectators attended a game of the highest division of the German league system. This aspect is particularly relevant for smaller municipalities, which attract thousands of spectators per game (Lindo *et al.*, 2018). Second, spectator violence may be the result of physiological arousal (Branscombe and Wann, 1992). Potential determinants of physiological arousal include increased heart-rate, blood pressure, and respiration, which may influence affect, cognition, and (anti-)social behavior. Third, Dollard *et al.* (1939) propose the idea of the frustration-aggression hypothesis.[7] If the favorite team suffers a defeat, this may precipitate aggressive fan behavior (see, for instance, the results of Card and Dahl (2011)). Fourth, Bandura (1973, 2007) postulates the notion of 'social learning' or mimetic behavior. Regardless of the outcome of the match, the mere observation of the game suffices to trigger violent actions.

Alcohol plays a crucial role in the context of spectator violence in German football stadiums. Cook and Durrance (2013) describes the pharmacological effects of alcohol consumption on aggression and cognitive functions. Alcohol consumption is associated with a loss of inhibition and impaired judgment. Furthermore, experiments have shown that participants exhibit more aggressive behavior after drinking. The 2018 edition of the Police Crime Statistics specifies that more than one in four assaults (26.2 percent) were committed under the influence of alcohol. While in some countries alcoholic beverages are prohibited on the premises (e.g. in Brazil since 2003), the rules in German stadiums are somewhat ambiguous. The DFB's security guidelines stipulate that the sale of alcoholic beverages is forbidden before and during games in the stadium. Nevertheless, with the approval of the responsible local security bodies, the hosting clubs can deviate from the regulations, on their own responsibility. Only in the case of high-risk games, the clubs are urged to comply with the ban.[8] The clubs, however, have a strong incentive to deviate from the ban as more than one-sixth (538 million Euros in the 2017/18 season) of the *Bundesliga* clubs' earnings are generated by matchday revenues (e.g. tickets and catering) and the sale of alcoholic beverages is a substantial part of this (Deloitte, 2019). For this reason, alcohol and its potential side effects are very present in German football arenas.

2.2.3 Previous Literature

The paper relates to previous literature that investigates the impact of large scale sporting events on various types of criminal behavior. Studies in the US American context often use offense reports from the National Incident Based Reporting System to investigate the impact of American (college) football games on crime. Rees and Schnepel (2009) exploit within agency variation to study the effects of Division I-A college American football games on various offense categories for the years 2000-2005.[9] They find an increase in assaults, vandalism, arrests for disorderly conduct, and alcohol-related offenses in hosting municipalities on game days. Larger effects are associated with unexpected game outcomes, defined as when lower ranked teams win against higher ranked teams. The effects of away games are not significantly different from zero. Lindo *et al.* (2018) examine the effect of college party culture in the context of Division 1 American football games on sexual

[7] Berkowitz (1989) provides a helpful overview of the frustration-aggression hypothesis.

[8] For details, please refer to: `https://www.sueddeutsche.de`.

[9] Almost all of the following studies exploit within law enforcement agency variation over time while controlling for weather, holidays and other sources of heterogeneity over time.

assaults. They show that the daily reports of rape victimization among 17-24-year-old women increase by 28 percent on games days. The effects are larger for home games, prominent matches, and games involving teams playing in the better ranked subdivision of Division 1. Furthermore, they show that game outcomes matter: unexpected wins lead to a strong increase in the number of rapes. Card and Dahl (2011) analyze the impact of emotionally unsettling events associated with wins and losses of professional American football teams on family violence for the years 1995-2006. They find a ten percent increase in intimate partner violence in the event of unexpected losses (when the home team was expected to win). There are no effects for unexpected wins or when the game expectations predict a close match. The effects of intimate partner violence of men against women are pronounced around the end of the game and are larger for higher-profile games.

In the European context, researchers focus on football matches to estimate the effect of large scale sporting events on crime. Marie (2016) investigates the effect of football matches on crime in London using hourly offense data from the Metropolitan Crime Statistics System. His results show that property crimes increase (decrease) by 4 percent (3 percent) for every additional 10,000 spectators attending a home (away) game. Violent crimes are only affected by derby matches. Montolio and Planells-Struse (2016) study the temporal impact of football matches on criminal behavior in Barcelona (2007-2011). They match reports of registered crime with football matches played by the Football Club Barcelona (FCB) to see whether the games lead to temporal shifts in criminal activity. They employ a panel approach comparing crime rates during the same time window on the same day of the week with and without FCB games. Their results indicate temporal shifts for criminal activities of thefts, criminal damage, robberies, and gender violence. Moreover, instances of gender violence increase after home defeats. In a follow-up study, Montolio and Planells-Struse (2019) investigate the spatial dimensions of crime externalities associated with football games. Their findings show that, in the event of a home game, theft rates (mainly pickpocketing) increase in the entire city. The impact is larger for regions in close proximity to the stadium. During away games, by contrast, the number of thefts occurring near the stadium is relatively smaller. The effects of football matches on assaults are analogous to thefts. In fact, on days with home matches, there are relatively more physical assaults in areas close to the stadium.

2.3 Data

The data set used for the analysis covers the time window from 2011 to 2015 and contains regions in which professional football games take place. The analysis is conducted at the municipality level, the smallest territorial division in Germany. I combine various data sources to examine the impact of professional football games on violent behavior.

2.3.1 Crime Data

The crime data is derived from the German Police Crime Statistics, which is provided by the Federal Criminal Police Office.[10] It includes the universe of individuals who were victim to

[10] Many aspects of the data preparation are inspired by Hener (2019) who uses the same data to examine the causal effect of noise pollution on criminal activities.

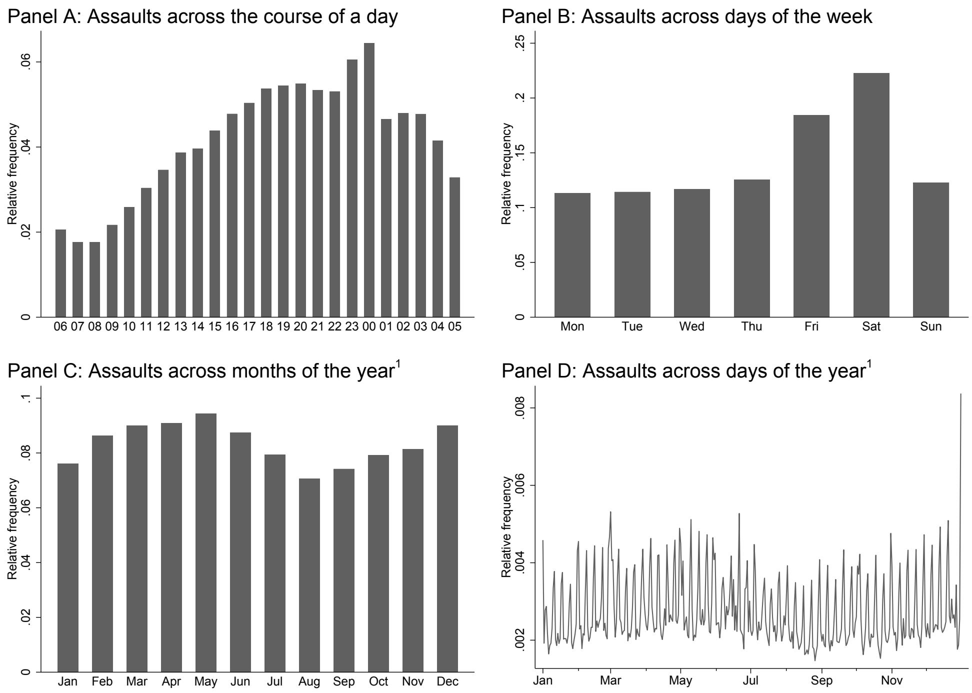


Figure 2.1. Distribution of assaults across time
Notes: The figure shows the distribution of assaults across hours of the day, days of the week, across months (adjusted for the number of days per month), and across the days of the year in the Federal Republic of Germany.

[1]: The figures in panel C and D are solely based on the year 2014. Please consult the appendix for figures from the other years.

a crime against their legally protected personal rights between 2011 and 2015. However, as the data is not reported until after police procedures are completed, only data from January 2011 to May 2015 is used to avoid problems with lags between the occurrence of the crime and the time of reporting. Besides, the month of June is excluded from the analysis as there are generally no matches during that time of the year. In addition to the time and place (municipality level) of the crime, the date include the crime type code, the victim's age and gender, information on how the attack was carried out (attempt/completed act, usage of a firearm, lone operator/crime was committed by a group) and information on the relationship between victim and suspect.[11] Roughly 40 percent of victims are female, the average age is 32 years, and 40 percent of the victims had no prior relationship with the suspect.

The micro-data is aggregated to the municipality-day level and the main outcome, the assault rate, is defined as the number of assaults per million population. Assaults are defined as actions involving physical violence. For that purpose, I use the crime type code 'simple willful bodily harm' (*Vorsätzliche einfache Körperverletzung*, § 223 StGB). There are roughly 120 types of criminal offenses (recorded in 6 digit codes), with the vast majority of cases classified by only a handful of codes.[12] Figure B.1 in the Appendix shows

[11]The relationship between victim and suspect is retrieved in two ways. On the one hand, formal relationships are recorded (such as types of kinship or acquaintance). On the other hand, relationships are defined in spatial-social terms (for instance living in the same household, or being in an educational or care relationship).

[12]The top 10 of the most prevalent crime keys account for more than 90 percent of the cases.

the distribution of cases per crime key for the twenty most common offense types in 2014. 'Simple willful bodily harm' is by far the most common offense, contributing to about 45 percent of all cases. Due to its prevalence and its association with aggression, I choose this offense type as the main outcome variable in the analysis.

Figure 2.1 shows the distribution of assaults over time. Panel A displays the variation of assaults per hour of the day.[13] The number of assaults increases during the day and peaks around midnight. To assign the cases that occur in the early morning hours to the day on which they originate, I define a day as beginning at 6:00AM and ending at 5:59AM. Panel B shows the distribution of assaults by day of the week. There are relatively more assaults on Fridays and Saturdays, whereas the other days exhibit slightly smaller assault rates. Panel C shows that the number of assaults has a strong seasonal pattern, with the highest value recorded in May and the smallest in August.[14] Panel D confirms this impression by plotting the daily number of assaults. New Year's Eve is a particularly impressive outlier.[15]

2.3.2 Football Data

The data on football matches is self-collected and is obtained via web scraping from www.kicker.de and www.transfermarkt.com. All matches played in the first three leagues of the German football league system in the period from January 2011 until May 2015 are recorded. The data contains detailed match and table standings parameters, e.g. time and place of the match, number of spectators, pregame point difference, goals, penalties, cards, referee characteristics, among others. Furthermore, there is comprehensive information on the individual teams, such as team size, average age, market value, and the number of foreign players. The stadiums where the matches take place are geographically encoded. Figure 2.2 depicts a map with all 69 stadiums included in the data set.

Figure 2.3 illustrates insights into key variables. Panel A shows the number of matches per day of the week and league. The vast majority of matches takes place between Friday and Sunday. Games of the lower leagues occasionally also take place during the week. Such games are held only in the evenings. In contrast, matches on weekends usually take place in the afternoon. The inclusion of day-of-week fixed effects in the baseline specification helps to account for the higher share of games played on weekends, which are associated with higher levels of criminal behavior. Spectator number vary substantially across the three professional leagues, as depicted in Panel C. The *Bundesliga* attracts the most spectators with an average of 44,000 viewers per game, followed by the second league with an average of 17,000 fans per match, and the lowest league attracts slightly less than 6,000 fans per game on average.

When investigating channels of how football games may affect assaults, I exploit betting odds obtained from www.oddsportal.com via web scraping. The betting odds give an idea of pregame expectations. I translate the odds of the three game outcomes to probabilities

[13] Roughly 15 percent of the observations do not contain hourly information. This has no consequences for the main analysis, as I examine daily variation in the assault rate.

[14] Panel C shows the monthly number of assaults while adjusting for the number of days per month.

[15] Panels C and D show data for the year 2014 only.

(a) The closest weather monitors

(b) Spatial displacement - neighboring regions

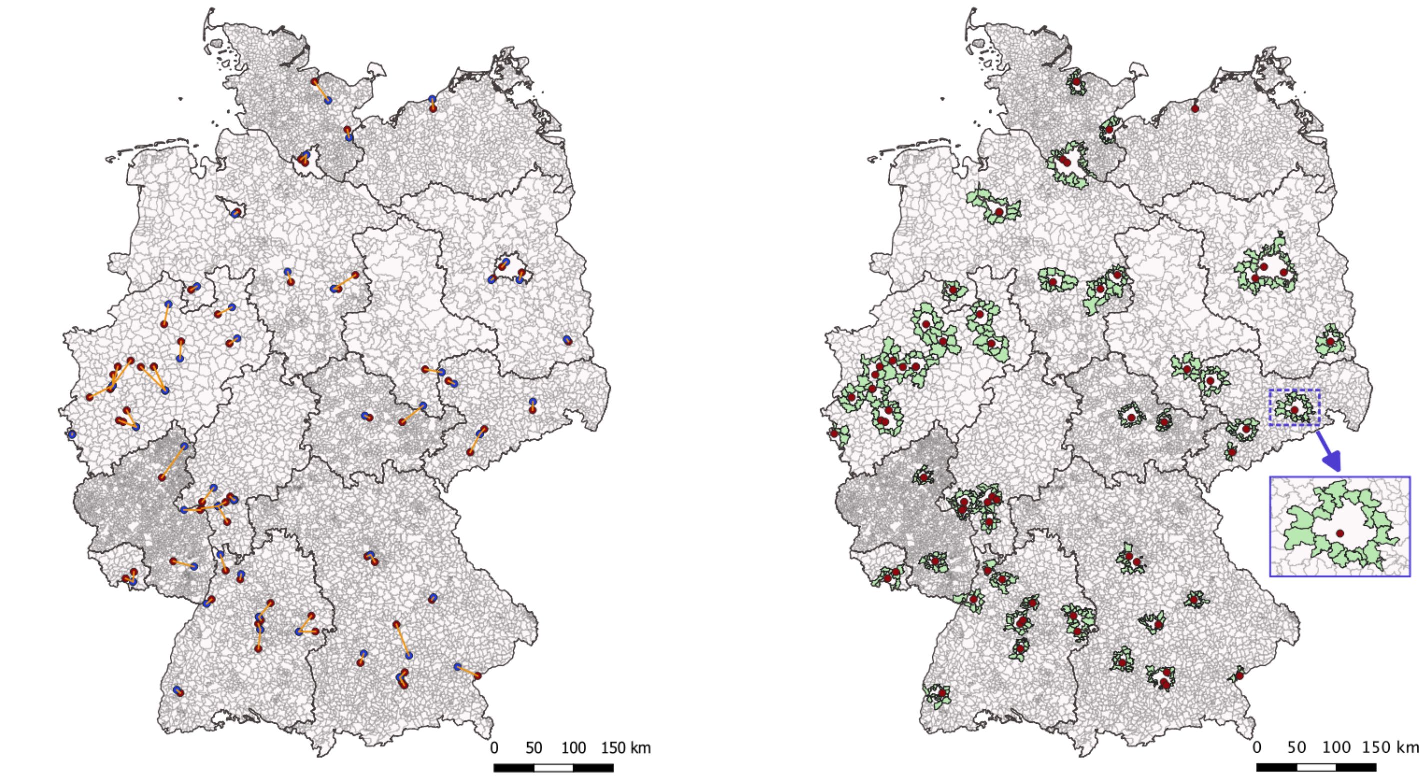


Figure 2.2. The stadiums with the closest weather monitors and neighboring regions
Notes: The map on the left shows the stadiums used in the analysis over the seasons 2010/11 until 2014/15 (red dots) and their closest weather monitors (blue dots). The orange lines indicate how the weather monitors are assigned to the stadiums. The map on the right shows the regions that are used in the analysis for spatial displacement effects. The neighboring municipalities are chosen to be in the sample for estimating spatial displacement effects if they have a common border with a region that contains a stadium. The red dots are the stadiums, the black outlines indicate federal state boundaries.
Source: Own representation with data from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).

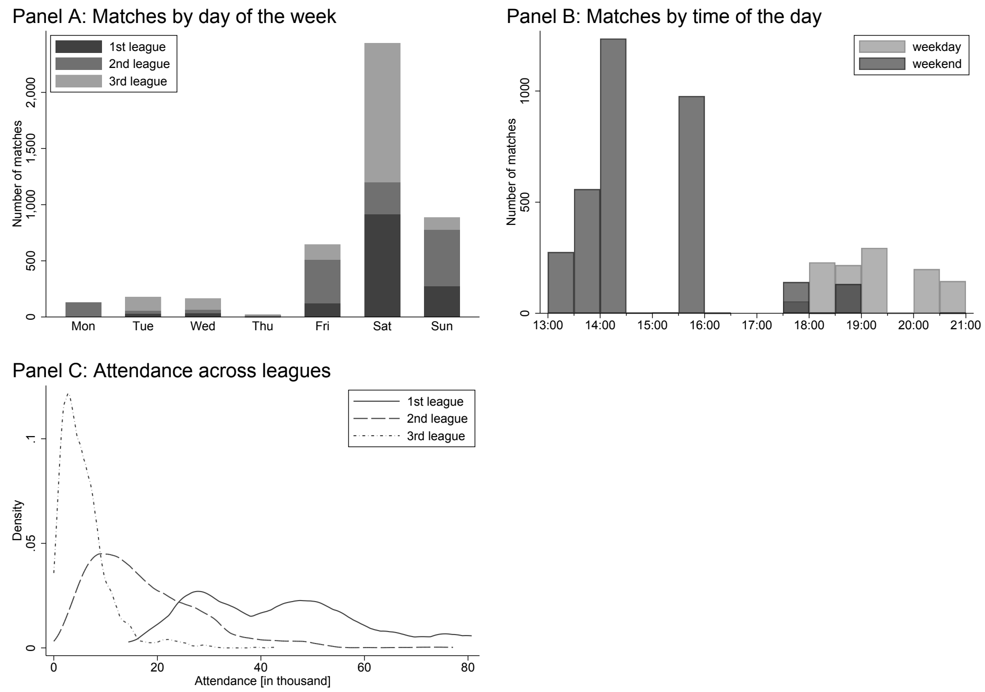


Figure 2.3. Football matches
Notes: The figures show key aspects of football games in the data set. Panel A shows how the number of matches vary over the course of a week, Panel B plots the distribution of matches over the course of a day, and Panel C shows kernel densities for the number of spectators (in thousand) across the three leagues.

which are the inverse of the betting odds. The probabilities serve as suitable predictors for game outcomes, as shown in Appendix Figure B.3.[16]

2.3.3 Weather Data

The weather data is derived from Germany's National Meteorological Service (*Deutscher Wetterdienst*). In order to construct the weather control variables, I use those weather monitors which measure the relevant weather variables in the sample period.[17] From this set of monitors, I choose the weather monitor with the closest proximity to a stadium. The assigned monitor-stadium pairs can be found in Figure 2.2. There is a high quality of the matches between weather monitors and stadiums, as the average distance between stadiums and monitors is 15 kilometers. Few of the weather variables have missing data, which are filled in by propagating forward from the last valid observation to the next valid observation (i.e. 'forward fill').[18]

[16] Panel A of Figure B.3 shows the close relationship between the realized score differential and the probability spread. Panels B and C demonstrate that the probability of winning increases the higher the probability spread.

[17] I use daily averages of the following weather variables: daily average, minimum and maximum air temperature, minimum ground temperature, vapor pressure, air pressure, cloud cover, air humidity, precipitation, hours of sunshine, snow depth, and wind velocity.

[18] The weather variables with missing data are (with the share of missing data in parenthesis): Cloud cover (<1.2 percent) and snow depth (<0.6 percent).

2.3.4 Holidays

In order to capture any variation in the crime rate between ordinary and special days, I add controls for public and school holidays, which may differ at the state level.[19] Furthermore, I add a dummy variable for peculiar days (New Year's Eve and the days surrounding Carnival), which are not holidays, yet certainly shift the crime rate.

2.3.5 Regional Database

The Federal Statistical Office and the statistical offices of the Länder provide a database of detailed statistics by various subject areas at a very granular spatial level. Thus, I am able to create a panel at the municipality-year level containing comprehensive information on topics such as territory, population, labor market, election results, housing, economic sectors, and public budgets. The information is used to construct weights for the regression analysis or to determine assault rates.

2.4 Empirical Strategy

In order to identify the causal effect of football matches on criminal behavior, I exploit within-region variation over time. To be precise, I compare the regional assault rate on a game day to the expected assault level conditional on the day of the week, month, and year, while additionally accounting for other possible confounding variation due to weather and holidays. In other words, the counterfactual regional assault rate (what would be expected on a game day in absence of the game), e.g. a Saturday in April 2012, is obtained by using the regional assault rate on other Saturdays in April 2012 with no games scheduled.

The identification strategy is based on a generalized difference-in-differences approach to study the impact of football matches on violent behavior. Let Assaults_{rdmy} denote the assault rate in region r, on day-of-the-week d, in month m and year y. The assault rate is defined as the number of assaults per million population and is given by:

$$\text{Assaults}_{rdmy} = \alpha + \beta\ (\text{Gameday}_{rdmy}) + \vartheta_r + \underbrace{\gamma_d + \eta_m + \theta_y}_{\text{date}_{dmy}} + \lambda \mathbf{X}_{rdmy} + \varepsilon_{rdmy}. \tag{2.1}$$

Gameday_{rdmy} is a binary variable that equals one when there is a home game, and zero otherwise. Region fixed effects ϑ_r capture time-invariant differences between regions and ensure that the identification is driven by within instead of between region variation over time. The vector date_{dmy} contains fixed effects for the day-of-the-week (γ_d), month (η_m), and year (θ_y). This way, the model flexibly controls for day-of-week specific heterogeneity, seasonal effects, and long-run time trends. I expand the baseline model by adding interactions of region fixed effects with all elements of date_{dmy}, i.e. region-by-day-of-week fixed effects, region-by-month fixed effects, and region-by-year fixed effects. The interactions account for systemic changes in the degree of violent behavior over the year for each region.

[19] The data on school holidays comes from 'The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany' (*Kultusminister Konferenz*). The data on public holidays is collected from https://www.schulferien.org/deutschland/feiertage/.

The vector $\mathbf{X}_{rdmy}$ includes school and public holidays and weather controls.[20] I compute two-way cluster-robust standard errors to capture arbitrary correlation at the region-year and year-month levels. Observations are weighted with the population figures from the Federal Statistical Office.

The implicit assumption for interpreting the parameter of interest β as the causal effect of a home game on violent behavior is that the location and the time of a football match are orthogonal to the number of assaults, conditional on the covariates. However, displacement effects may pose a threat to identification. On the one hand, this refers to spatial displacement effects, which may occur when (violence-prone) people from distant regions visit a game. On the other hand, this includes temporal displacement effects, which happen when assaults are shifted from adjacent days to game days. In both cases, the parameter would overestimate the impact of a football match on violent behavior as the offense would have been committed regardless, but at a different time or place. To rule out the possibility that displacement effects compromise the validity of the identification strategy, I investigate the effect of football games on neighboring regions and on days adjacent to game days in section 2.5.2. I find that the main results, namely an increase in violent behavior in regions where football games take place, are not neutralized by a decrease in the number of assaults in surrounding regions or on days adjacent to game days.

Given the design of the empirical approach, there could be another potential threat to the validity of the identification strategy. By focusing on home games in the main analysis, the counterfactuals may be biased downwards as days with away games are part of the control group. This control group problem may be due to violent fan groups traveling with their team to away games, potentially leading to a decline in the assault rate in the home region. To address this concern, I perform the analysis again, differentiating between home and away games. When considering distinct effects for home and away games, I find that away games do not significantly affect the assault rate in the home region.

2.5 Results

2.5.1 Main Results

Before presenting the regression results, Figure 2.4 gives an intuitive preview of the main findings. Using the same data as in the main analysis, it shows the assault rate across the days of the week.[21] The daily rates are presented for days with and without home games. The average daily assault rate is higher for weeks when a home game is played than when no game is played. The difference in means is statistically significant for all days except

[20]Public holiday controls include binary variables (at the level of the Federal States) for All Saints' Day, Ascension Day, Assumption Day, Christmas, Corpus Christi, Epiphany, Easter, German Unity Day, Good Friday, Labor Day, New Year's Day, Penance Day, Pentecost, and Reformation Day. Moreover, it contains dummy variables for Carnival and New Year's Eve.
Weather controls (at the regional level) include average air temperature, maximum air temperature, minimum air temperature, minimum ground temperature, steam pressure, cloud cover, air pressure, humidity, average precipitation, hours of sunshine, snow depth, and wind speed.

[21]Appendix Figure B.4 shows the same scheme using the raw number of assaults across days of the week for days with and without games.

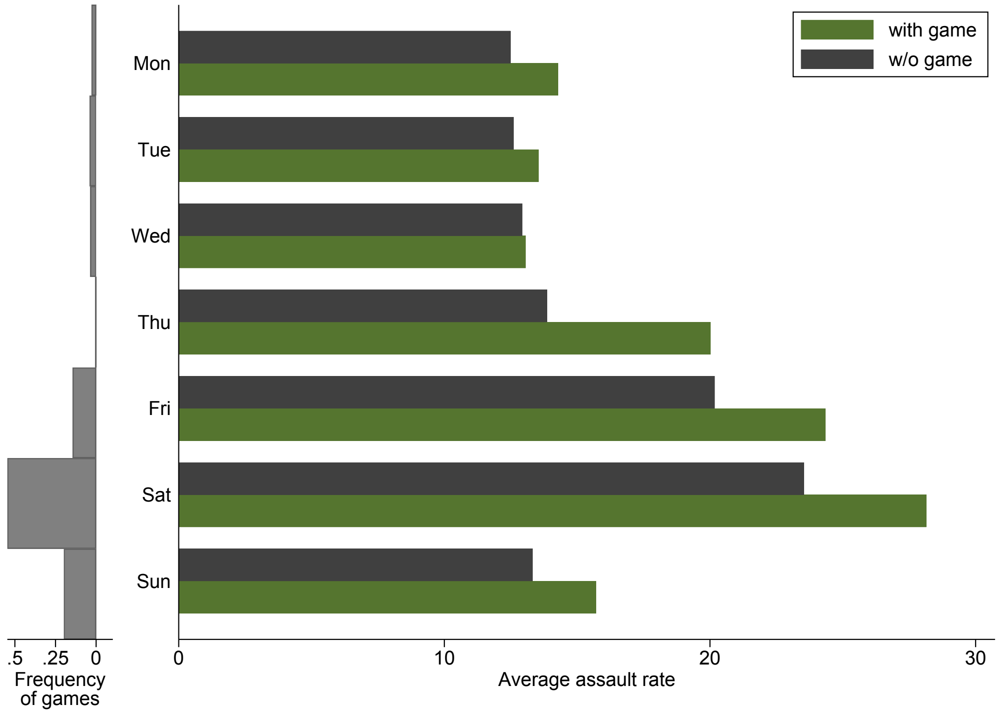


Figure 2.4. The average assault rate on gamedays and days where no game takes place
Notes: The figure shows the daily average number of assaults per million population for regions that host games of a football team from the top three leagues. The daily rates are shown for weeks in which a game is played and for weeks in which no game takes place.

Tuesdays and Wednesdays.[22] The empirical model exploits the variation in the number of assaults across the days of the week, and in particular how the pattern varies between weeks with and without games.

Table 2.1 reports estimates corresponding to equation 2.1 when continuously adding more controls. The dependent variable is the assault rate, which is defined as the number of offenses per million population. In column 1, I include region, day-of-week, month, and year fixed effects. In column 2 weather controls are added. Holidays fixed effects are included in column 3. In column 4, I control for region-specific date fixed effects. Note that although the estimates vary marginally across columns, I use throughout the paper the model presented in column 4 for the analyses that follow. The estimate from the preferred specification in column 4 suggests that a home game increases the assault rate by an average of 2.677 assaults per million population. This corresponds to an increase of 21.5 percent compared to the level without games.

Having established that football games lead to more assaults, I now shed light on the victims of these additional offenses. Panel A of Table 2.2 shows the effect of football games on the assault rate by victim characteristics. First, columns 1 and 2 show the effect heterogeneity by gender. Although the estimates for women and men are statistically different from zero, the vast majority of additional victims are male. Male victimization rates increase on average by 2.432 offenses per million population. The increased assault

[22]The average assault rate on Thursdays with games is very high compared to Thursdays without games. This may be the result of very few games take place on a Thursday. In my sample, only 18 (0.4 percent) of all games take place on a Thursday.

Table 2.1. Effects on assault rate

	(1)	(2)	(3)	(4)
Game day	2.740***	2.766***	2.813***	2.677***
	(0.319)	(0.312)	(0.313)	(0.284)
Effect size [%]	22.00	22.21	22.59	21.50
Observations	88,028	88,028	88,028	88,028
Region FE	✓	✓	✓	✓
Date FE	✓	✓	✓	✓
Weather Controls	-	✓	✓	✓
Holiday FE	-	-	✓	✓
Interact FE	-	-	-	✓

Notes: Estimates are based on the model shown in equation 2.1. The specifications use daily data (excluding June) spanning the time window 2011-2015 for regions that host games of a football team from the top three leagues of the German football league system. The outcome variable is defined as the number of assaults per million population. Population-weighted coefficients show the change in the outcome variable due to a home game. Days are defined to run from 6:00AM until 5:59AM the following day to accommodate the fact that offenses committed in the early morning hours have their origin in the preceding day. The effect size corresponds to the percent change of the assault rate due to a football game in relation to the mean when no game takes place. Control variables shown as *Date FE* include dummies for day-of-week, month, and year. *Weather controls* include air temperature (average, maximum, and minimum), minimum ground temperature, vapor pressure, air pressure, cloud cover, air humidity, precipitation, hours of sunshine, snow depth and wind velocity. *Holiday FE* are dummy variables for public and school holidays, as well as for other peculiar days. Control variables shown as *Interact FE* consist of interactions of region dummies with all elements of the date fixed effects. Two-way clustered standard errors at region-year and year-month level are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

rate for males accounts for more than 90 percent of the effect found for the entire sample. Figure 2.5 shows the age profile of the impact of football games on the assault rate for each gender. For women, the point estimates are small in magnitude and not significantly different from zero. Women aged 40-49 are the only exception. In contrast, the effects for adult men are throughout significant. The largest effect for males is found in the 18-29 age group and decreases thereafter. Second, I explore effect heterogeneity according to the relationship between victim and suspect. Columns 3 and 4 distinguish the relationships from a formal perspective, such as kinship or acquaintance. Although both estimates are positive and statistically significant, the majority of additional assaults involves victims with no prior connection to the suspect. The victimization rate of strangers to the suspect increases by on average 1.939 assaults per million population. This implies that almost three out of four additional cases involve this type of victim-suspect pairing. Column 5 considers spatial-social relationships, namely whether victim and suspect live in the same household. A football game increases the number of domestic assaults by on average 0.096 assaults per million population. This implies that only few (13 percent) victims who know the suspects also live in the same household with them. Third, the effect of the victim's occupation is investigated. Column 6 shows the impact of football games on violent behavior directed at police officers on duty. The victimization rate of police officers increases on average by 0.434 assaults per million population. Thus, almost 16 percent of the additional assaults resulting from football games are attributed to attacks on police officers. In short, the additional victims are young males with no prior relationship to the

Table 2.2. Effect heterogeneity by victim and crime characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Victim characteristics						
	Gender		Victim-suspect-relationship			Occupation
	Women	Men	Strangers[a]	Prior[a] relation	Domestic[b]	Police
Game day	0.245***	2.432***	1.939***	0.738***	0.096**	0.434***
	(0.090)	(0.232)	(0.207)	(0.116)	(0.045)	(0.078)
Effect size [%]	4.99	32.27	40.43	9.64	5.40	96.98
Observations	88,028	88,028	88,028	88,028	88,028	88,028
Panel B: Crime characteristics						
	Timing[c]				Assault execution	
	Spring	Summer	Fall	Winter	Attempt	Completed act
Game day	2.817***	1.482**	2.478***	2.849***	0.367***	2.310***
	(0.404)	(0.484)	(0.372)	(0.535)	(0.060)	(0.238)
Effect size [%]	21.00	13.40	20.71	22.51	56.21	19.58
Observations	27,140	14,632	21,476	24,780	88,028	88,028

Notes: The estimates are based on the model shown in equation 2.1 and use the same outcome and controls as column 4 of Table 2.1 (including region and date fixed effects, their interactions as well as holiday and weather controls). See Table 2.1 for additional details. Two-way clustered standard errors at region-year and year-month level are reported in parentheses.
Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.
[a]: covers formal relationships (e.g. types of kinship or acquaintance).
[b]: covers spatial-social relationships (whether victim and suspect live in the same household).
[c]: spring (Mar-May), summer (Jul+Aug), fall (Sep-Nov), winter (Dec-Feb).

suspect. Hence, the victims and the people who usually gather in and around football stadiums are similar in demographic terms (PwC, 2016).

In the next step, I consider effect heterogeneity by crime characteristics. Columns 1-4 in Panel B of Table 2.2 show the heterogeneity of the effect by the time of the offense. The impact of a football game on the assault rate is smallest in summer months, with 1.482 assaults per million population, representing an increase of 13.4 percent. The football season starts in the summer months (July and August). At this point, the results of the games are not yet so important. The estimates are higher (but not significantly different from each other) at other times of the season. Columns 5 and 6 consider effect heterogeneity by assault execution (attempt/completed offense). I find that the effects of football games on the assault rate are mostly driven by completed offenses with on average 2.310 additional assaults per million population. The share of attempted assaults also increases, but at a smaller rate of 0.367 additional attempted assaults per million population on average.

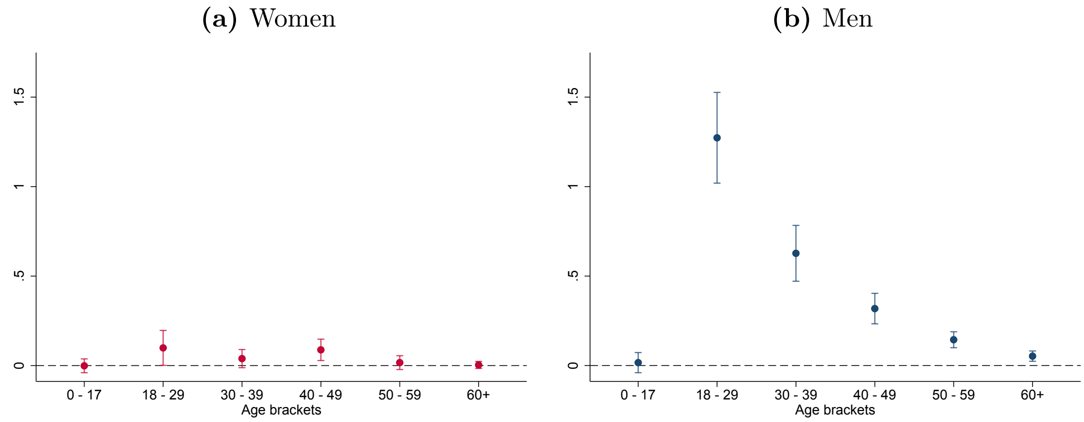


Figure 2.5. The age profile of the impact of football matches on the assault rate
Notes: The figure shows estimates and 95% confidence intervals across age brackets and by gender. To estimate the effect, I use the model shown in equation 2.1 in its richest specification with all covariates and population weights. The outcome variable is defined as the number of assaults per million population.

2.5.2 Potential Threats to Identification and Validity of the Design

In this section, I consider the possibility that my results do not reflect additional assaults due to football games, but merely shifts in offenses. Furthermore, I test the sensitivity of the main results by additionally including away games in the model.

Previously, I have presented evidence that football games increase the assault rate. However, it is possible that I only capture an effect that shifts offenses. For example, the increase in violent behavior may be offset by a decline in assaults in other areas or at different times (Lindo *et al.*, 2018). In other words, the assault would have been committed regardless, but at a different time or location. One explanation may be different population flows around days on which games take place.

In order to estimate *spatial displacement effects*, I investigate the impact of football matches on neighboring regions. A neighboring region is a municipality that shares a border with a region in which a stadium is located.[23] Figure 2.2 shows a map of the selected regions. The sample of neighboring regions exhibits a considerably higher number of observations. This is owed to the fact that a region with a stadium has on average slightly more than 11 neighboring municipalities. Panel A of Table 2.3 shows the estimates of the impact of a home game on these neighboring regions. In comparison to the baseline effects, the spatial spillover coefficients are small and not significantly different from zero. Consequently, the results do not suggest offsetting spatial spillover effects.

In the next step, *temporal displacement effects* are considered. To capture these effects, I include a one-day lead and lag of the game day indicator. Panel B of Table 2.3 contains the estimates when including the temporal spillover components in the baseline model. The estimates of the game day itself are not significantly different from the baseline model. Almost all of the coefficients for the day before and after the game are small in magnitude and not statistically significant. The only exception is the estimate for the day before the game in column 4, which implies that a football game leads to an average increase of 0.352 assaults per million population on the day before a game. The positive coefficient,

[23]If two municipalities share a border and each of the regions contains a stadium, both regions will not serve as neighbor regions and they are dismissed from the set of spatial spillover candidates.

Table 2.3. Displacement effects

	(1)	(2)	(3)	(4)
Panel A: Spatial displacement				
Game day	0.144	0.153	0.170	0.210
	(0.146)	(0.146)	(0.147)	(0.154)
Effect size [%]	2.58	2.74	3.04	3.75
Observations	960,848	960,848	960,848	960,848
Panel B: Temporal displacement				
Game day	2.777***	2.810***	2.857***	2.770***
	(0.346)	(0.338)	(0.336)	(0.300)
Day after game	-0.105	-0.076	-0.065	0.279
	(0.189)	(0.190)	(0.192)	(0.182)
Day before game	0.287	0.318	0.339	0.352*
	(0.241)	(0.238)	(0.229)	(0.206)
Effect size [%]	22.34	22.60	22.98	22.28
Observations	87,438	87,438	87,438	87,438
Region FE	✓	✓	✓	✓
Date FE	✓	✓	✓	✓
Weather Controls	-	✓	✓	✓
Holiday FE	-	-	✓	✓
Interact FE	-	-	-	✓

Notes: Estimates are based on the model shown in equation 2.1. Panel A contains specifications that use daily data (excluding June) spanning the time window 2011-2015 for regions that share a border with a distract in which a stadium is located. Panel B shows specifications that use daily data (excluding June) spanning the time window 2011-2015 for regions that host games of a football team from the top three leagues. The outcome variable is defined as the number of assaults per million population. Population-weighted coefficients show the change in the outcome variable due to a home game. Days are defined to run from 6:00AM until 5:59AM the following day to accommodate the fact that offenses committed in the early morning hours have their origin in the preceding day. The effect size corresponds to the percent change of the assault rate due to a football game in relation to the mean when no game takes place. The estimates are based on the model shown in equation 2.1 and use the same set of controls as column 4 of Table 2.1 (including region and date fixed effects, their interactions as well as holiday and weather controls). See Table 2.1 for additional details. Two-way clustered standard errors at region-year and year-month level are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

however, does not suggest temporal displacement effects, which would require a negative estimate. Rather, it indicates additional assaults due to activity on the day preceding important games.

As discussed above, the baseline model considers the effect of football matches, but only for home games. This restriction may compromise the validity of the design. When away matches are not accounted for, they end up in the control group. The control observations might be biased downwards if the most devoted (and possibly violent) fans leave their home municipality to accompany their local team to an away game. The resulting decrease in the assault rate at home due to the absence of local agitators imply that days with away matches can no longer function as control units. To address this concern, I investigate

Table 2.4. Effect of football games on the assault rate, distinction of home and away games

	(1)	(2)
	Baseline w/o L3	Distinction home/away[1]
Game day	2.859***	
	(0.324)	
Home game day		2.910***
		(0.335)
Away game day		0.325
		(0.213)
Effect size [%]	21.22	21.97
Observations	61,172	61,172

Notes: The specifications use daily data (excluding June) spanning the time window 2011-2015 for regions that host games of a football team from the top two leagues. The estimates are based on the model shown in equation 2.1 and use the same set of controls as column 4 of Table 2.1 (including region and date fixed effects, their interactions as well as holiday and weather controls). See Table 2.1 for additional details. Two-way clustered standard errors at region-year and year-month level are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[1]: Effect size corresponds to the coefficient of Home game day.

the *effect of home and away matches separately.* To analyze the effect of away matches, the design of my data set must be modified. In the baseline version, the football data is merged with other datasets at the match level (the region ID of the home team serves as the identifier). In this case, I use the football data at the table standings level. In other words, both the home and away teams are matched with a region. This approach leads to ambiguity regarding the treatment status of individual regions.[24] For instance, the treatment status of regions with more than one team is ambiguous when there is a home and an away match on the same day. To alleviate this concern, I exclude the third league from the sample and focus exclusively on the first two leagues.[25] This approach helps considerably to clarify the treatment status of a region. Table 2.4 shows the results when home and away matches are examined separately. To compare the estimated effects, column 1 shows estimates retrieved from the baseline model (home matches only) when the sample is adjusted as described above. Column 2 presents the estimates that incorporate the impact of home and away games on the assault rate. The effect of a home game is sizable, leading to an average increase in the assault rate of 2.910 assaults per million population. The coefficient is not significantly different from that in column 1. A negative and significant estimate of an away game would compromise the identification strategy. However, the estimate of an away game is small in magnitude and not significantly different from zero. Thus, the results suggest that focusing exclusively on home games does not render the identification strategy invalid.

[24]When only considering home games, the treatment status is not a problem. This is due to the fact that local authorities do not allow two home games on the same day.

[25]Some ambiguities remain, but they are solved as follows: 12 percent of the matches still include a duplication of two teams per region playing on the same day, either one home and one away game, or two away games. In the latter case, the status of the region is defined as 'away'. In the former case, it is defined as 'home'.

2.5.3 Channels

This section investigates potential mechanisms through which football games may cause additional assaults. I consider emotional cues and the prominence of games as potential channels.

First, I consider emotional cues. This is motivated by the findings of Card and Dahl (2011) who demonstrate that unexpected defeats of local football teams trigger family violence. The results can be best explained with the frustration-aggression hypothesis, first proposed by Dollard *et al.* (1939), which predicts aggressive behavior in the event of frustrating events. Rees and Schnepel (2009) similarly show that there are more violent offenses when the local college football team suffers a defeat. For this reason, I analyze in Table 2.5 whether visceral factors may be the reason for the additional assaults. In column 1 I investigate whether an emotionally upsetting event during a game leads to a higher assault rate. To answer this question, I create an index that equals one for games that include at least one of the following potentially troubling events: a penalty is awarded (20 percent of all games), a player receives a red card (10 percent of all games), or the referee receives an insufficient grade (15 percent of all games). The index shows that 35 percent of all games involve at least one upsetting episode as defined in the previous categories. The estimates in column 1 do not suggest that emotional cues trigger more violent behavior since the estimates for games with and without upsetting events are not significantly different. In the second and third column, I show estimates following the approach of Card and Dahl (2011). I examine the impact of game outcomes relative to their pregame expectations. Pregame expectations are included in the analysis as matches with contrasting predictions may be very different from each other. By including predicted outcomes, I can estimate the effect that results from the defeat of a team that was expected to win, and vice versa. Using data from oddsportal.com, I define a game as unpredictable when the absolute probability difference between winning and loosing is smaller than 20 percentage points.[26] When the spread's value exceeds the threshold, a win or a loss of the home game is expected. Around 45 percent of the games are expected to be close, another 45 percent are expected to be won, while 10 percent of the games are expected to be lost. The significantly larger share of expected victories may be attributed to the home-advantage. In column 2, I first examine the effects of matches with distinct predicted match outcomes. The estimates do not suggest that the effect of games with different predicted outcomes vary systematically from each other. In column 3, I additionally include interactions between expected and actual game outcomes. The estimates are relatively small in magnitude and not significantly different from zero, implying that unexpected wins/losses do not cause additional assaults. Altogether, there is no evidence that emotional cues drive violent behavior in the context of professional football games in Germany.

Second, I examine game and team prominence as potential channels. For many sports fans, the affection for their football team plays a central role (Wann and Branscombe, 1993). Almost as important is the cultivation of animosities against rivals. Resentment and hatred is likely to lead to violent actions (Nassauer, 2011). To test the hypothesis that the additional assaults are the result of hostile feelings, I compare the impact of matches played between known rival teams to regular matches. The game day indicator from equation 2.1 is replaced by an interaction with a dummy variable that equals one for high-rivalry matches. Local derbies (games between two competing teams that are based in

[26]Similar results are obtained when I define different threshold values and when I deviate from the symmetry around the origin.

Table 2.5. Effect of emotional cues

	(1)	(2)	(3)
		Card & Dahl (2011) specification	
	Upset event index	Predicted outcomes	Predicted and actual outcomes
Upset event (Index)	2.717***		
	(0.406)		
No upset event (Index)	2.655***		
	(0.289)		
Expected to lose		3.495***	3.376***
		(0.722)	(0.836)
Expected to win		2.752***	2.907***
		(0.366)	(0.356)
Expected to be close		2.437***	2.369***
		(0.321)	(0.383)
Expected to lose and won			0.599
			(1.554)
Expected to be close and lost			0.194
			(0.610)
Expected to win and lost			-0.792
			(0.667)
Observations	88,028	88,028	88,028

Notes: The gameday indicator is replaced by an index that captures unsettling events. The upset event index in column 1 is defined as a dummy variable equal to one if one of the following events take place: a penalty is awarded (20% of all games), a red card is being issued (10% of all games), or the referee receives a non-sufficient grade (15% of all games). In columns 2 and 3, I use data from oddsportal.com to classify games as expected to win/lose/be close. The estimates are based on the model shown in equation 2.1 and use the same set of controls as column 4 of Table 2.1 (including region and date fixed effects, their interactions as well as holiday and weather controls). See Table 2.1 for additional details. Two-way clustered standard errors at region-year and year-month level are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

regions of close geographical proximity) constitute high-rivalry matches.[27] The estimates are shown in columns 1 and 2 of Table 2.6. Games that are classified as high-rivalry matches lead to an average increase of 8.320 assaults per million population. This corresponds to more than three times the baseline effect. Although the standard errors are relatively large, considering that only 2.5 percent of the games are classified as high-rivalry matches, the effect is significantly different from that for regular matches. The higher rate of physical assaults during high-rivalry matches could also be rationalized by the social learning theory, according to which spectators mimic behavior seen on the field. For instance, the number of yellow and red cards is significantly higher in high-rivalry matches.[28] Finally, the mere prominence of teams/games might attract more (violent) people. As shown in Figure 2.3, the number of spectators varies substantially in the various divisions of the German football league system. Columns 3-5 shows the effect of games by the league in which the

[27] Appendix Table B.2 gives an overview of high-rivalry matches.

[28] The number of cards in derby matches is 4.324 (s.e.=0.174), while the number in regular matches is 3.906 (s.e.=0.029).

Table 2.6. Effect of game and team prominence

	(1)	(2)	(3)	(4)	(5)
	Rivals/derbies		League		
	High-rivalry	Regular	League 1	League 2	League 3
Game day	8.320***	2.387***	3.530***	1.878***	1.569***
	(1.746)	(0.275)	(0.460)	(0.318)	(0.350)
Effect size [%]	66.81	19.17	28.35	15.08	12.60
Observations	88,028	88,028	88,028	88,028	88,028

Notes: The estimates are based on the model shown in equation 2.1 and use the same set of controls as column 4 of Table 2.1 (including region and date fixed effects, their interactions as well as holiday and weather controls). See Table 2.1 for additional details. The gameday indicator is replaced by interactions with dummy variables for rival matches and league affiliation. Two-way clustered standard errors at region-year and year-month level are reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

teams play. The largest effect can be found for the highest league, the *Bundesliga*, in which one game leads to an increase in the assault rate by on average 3.530 assaults per million population. The impact of first division matches is significantly larger than the effect of matches in the other two leagues. Summing up, there is evidence that the prominence of games is a crucial determinant of the effect size of football matches on the assault rate, with high-rivalry matches and top league games leading to more assaults.

2.5.4 Robustness Tests

I perform several sensitivity and placebo tests to assess the robustness of the findings. The results of these checks are reported in Table 2.7. Overall, the sensitivity tests demonstrate that the main results are robust to alternative specifications and estimations, indicating that football games do indeed lead to more assaults.

Alternative Econometric Specifications.—First, I show that the results are not sensitive to alterations in the *sample*. Adjustments to the sample may be necessary as some of the games are played on different days than originally planned. Deviation from the original match schedule may pose a risk to the allocation of games that is plausibly random. For this reason, I exclude the set of rescheduled games from the analysis. The results are almost identical to the baseline results.[29] Second, I show the effects of omitting population *weights* from the regressions. The effect of football matches without considering population weights is considerably larger: a home game leads to an increase in the average assault rate by 5.342 assaults per million population. The higher coefficient compared to the baseline specification indicates that the effects of rural areas must be larger, as they become more relevant when population weights are omitted. Third, I show that my results are robust to a different *estimation* procedure and an alternative measure of the *dependent variable*. In the following, I use the raw number of assaults as the outcome variable. Given the discrete

[29] Of the 4,461 games in the sample, 2.24 percent are rescheduled. The vast majority of rescheduled matches (95 percent) take place in league three.

Table 2.7. Robustness tests: Impact on assault rate

	(1) Coefficient	(2) Standard error	(3) Effect size [%]	(4) N
Baseline	2.677***	(0.284)	21.50	88,028
Econometric specification				
Drop delayed games	2.699***	(0.288)	21.67	87,928
No population weights	5.342***	(0.566)	42.90	88,028
Poisson model[1]	1.483***	(0.238)	28.55	87,475
Other forms of violence				
Broadly defined assaults	5.417***	(0.529)	27.76	88,028
Threats	0.216***	(0.079)	5.44	88,028
Resistance to enforcement	0.775***	(0.103)	45.50	88,028

Notes: The specifications use daily data (excluding June) spanning the time window 2011-2015 for regions that host games of a football team from the top three leagues. The outcome variable is defined as the number of assaults per million population. Except where otherwise noted, the specifications use population-weights. Days are defined to run from 6:00AM until 5:59AM the following day to accommodate the fact that offenses committed in the early morning hours have their origin in the preceding day. The effect size corresponds to the percent change of the assault rate due to a football game in relation to the mean when no game takes place. All specifications use region and date fe, their interactions, weather controls, and holiday FE. Two-way clustered standard errors at region-year and year-month level are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[1]: number of assaults is dependent variable.

nature of the dependent variable and the fact that there are many cells with zero reports, I employ a Poisson model. The corresponding regression specification reads as follows:

$$\begin{aligned} E\left[\text{Assaults}_{rdmy} \,|\, \text{Gameday}_{rdmy}, \vartheta_r, \text{date}_{dmy}, \mathbf{X}_{rdmy}\right] \\ = exp\left(\beta\ (\text{Gameday}_{rdmy}) + \vartheta_r + \text{date}_{dmy} + \lambda \mathbf{X}_{rdmy}\right). \end{aligned} \tag{2.2}$$

Due to the nonlinearity of the model, the coefficient in Table 2.7 shows average marginal effects of a home game on the number of physical assaults.[30] A home game is predicted to lead to an average increase of 1.483 assaults. This represents a 28.55 percent increase in the number of assaults.

Other Forms of Violence.—Next, I investigate the robustness of the findings when considering at other forms of violence. First, a broader definition of assaults is considered. In addition to offenses coded as 'simple willful bodily harm', I further include negligent, dangerous, and grievous bodily harm as well as brawls. Appendix Table B.1 illustrates which offenses are included in the expanded definition of assaults and how the penal codes of the German Criminal Code (*StGB*) are translated into the offense keys of the Police Crime Statistics. The effect size of 5.417 assaults per million population is significantly larger than the baseline coefficient. This indicates that football games also lead to an increase in other forms of physical violence. Second, I investigate the effect of football games on

[30] In nonlinear models, coefficients cannot be interpreted as marginal effects. The partial effect for a Poisson model is given by $\frac{\partial E[\text{Assaults}|\mathbf{X}]}{\partial x_j} = \beta_j\ exp(\mathbf{x}'\beta)$. In order to present a single response value, I follow Cameron and Trivedi (2005) and report the average response: $\frac{1}{N}\sum_i \frac{\partial E[\text{Assaults}_i|\mathbf{X}_i]}{\partial x_{ij}} = \hat{\beta}_j \times \frac{1}{N}\sum_i\ exp(\mathbf{x}_i'\hat{\beta})$.

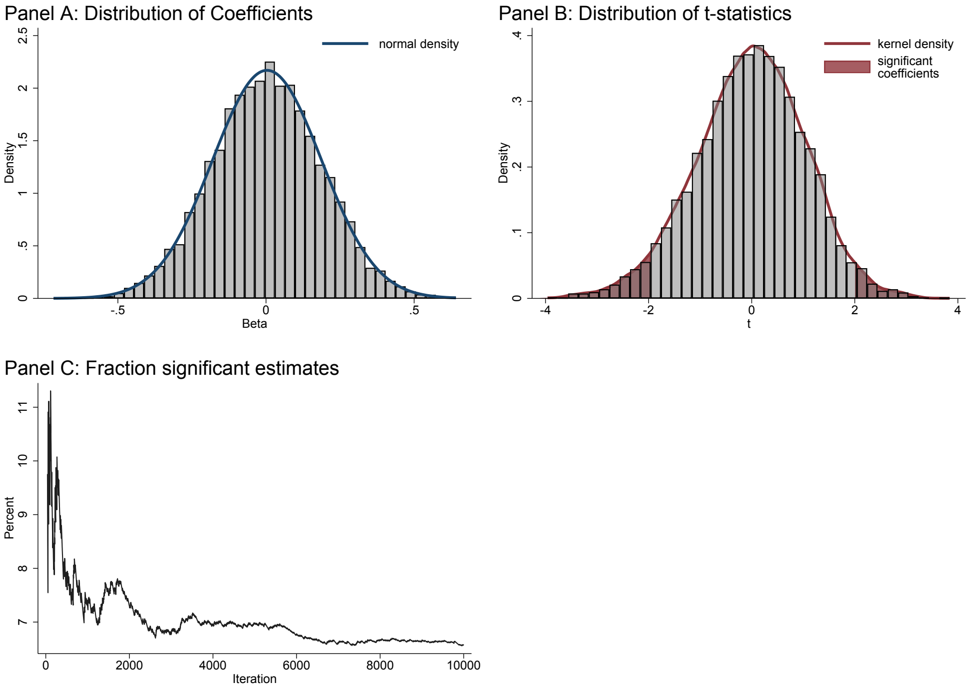


Figure 2.6. The effect of placebo games
Notes: The figure shows the effect of placebo games. Panel A presents the distribution of the coefficients (along with a normal density) after 10,000 iterations. Panel B shows the distribution of the t-statistics and the resulting ranges of significant coefficients, with a level of significance $\alpha = 0.05$. Panel C shows the fraction of significant estimates across the number of iterations.

threats, which can be regarded as a precursor of physical violence. The effect size of 0.216 threats per million population is relatively small. The estimate represents a 5.4 percent increase. Third, offenses coded as resisting law enforcement officers are examined. I find that football games lead to on average 0.775 offenses per million population. Overall, the results of these robustness checks indicate that other forms of violence are also affected by professional football games. This suggests that there is no substitution of different types of crime, but that additional violent offenses are committed.

Placebo Games.—I estimate the impact of *placebo games* on the assault rate to test whether the previous results are only due to chance.[31] The actual matches take place on about five percent of the days in the sample. To estimate the effect of placebo games, I drop the affected days with the actual matches and randomly assign dummy indicators with the same frequency of the real matches. Subsequently, I estimate the model as shown in equation 2.1. This procedure is carried out 10,000 times and the results are shown in Figure 2.6. Panel A displays the distribution of the coefficients. As expected, the coefficients are centered around zero. Panel B illustrates the distribution of the t-statistics. The red area below the kernel density indicates significant estimates for a significance level of $\alpha = 0.05$. Note that the t-statistics of the preferred specification in Table 2.1 (2.677/0.284=9.43) is almost 2.5 times higher than the largest observed value in any of the 10,000 simulations. Panel C shows that with 10,000 iterations, 6.57 percent of the estimates are significantly

[31] Unfortunately, there are no offenses that can function as placebo outcomes. This is because most of the offenses covered in the PCS are potentially affected by football games.

different from zero. At a significance level of $\alpha = 0.01$, there are 3.06 percent significant estimates. The low levels of significant coefficients confirm that the previous results are not due to chance.

2.6 Discussion and Conclusion

In this paper, I analyze the impact of professional football games on violent behavior. To estimate the causal effects of football games on physical assaults, I use a generalized difference-in-differences approach that exploits variation in the timing of matches. I compare regional assault rates on days with and without matches, conditional on the day of the week, month, and year while additionally accounting for potential confounding variation coming from weather and holidays. I match web-scraped information on 4,461 football matches with data on local assault rates, weather, holidays, and population figures to construct a panel at the municipality-day level for the period 2011-2015. I find that a home game increases the average assault rate by 21.5 percent. Male victimization rates drive the results and are particularly high for the 18-29 age group. Besides, the the effects are larger for victims with no prior relationship to the suspect and for completed offenses. There are no offsetting reductions in assault rates on days adjacent to game days or in nearby regions. In examining potential channels, I find no evidence that emotional cues are responsible for the increase in assaults on game days. Although there is no evidence supporting the frustration-aggression hypothesis, I find large effects for prominent games and teams. In fact, this set of results suggests that spectators mimic player behavior, select into specific matches, or that the mere agglomeration of fans leads to an increase in the assault rates on game days.

I find the external effects of football games on violent behavior to be large in magnitude and economically relevant. For instance, professional football games explain 17.7 percent of all assault reports in the regions in which professional football clubs are located. Back-of-the-envelope calculations indicate that football games in the top three leagues of the German league system precipitate an additional 18,770 assaults in the 2014/15 season,[32] which translate into annual social cost of 95 million euros.[33] The Coase (1960) theorem states that market failures resulting from externalities can be solved without public intervention (taxpayers do not have to pay for police operations) under the following conditions: clearly defined property rights, complete information, and low/no transaction costs. The question is whether the negative externalities would be internalized in absence of state action. To this end, the football clubs must (i) have an incentive to prevent violence at the events and (ii) be able to implement the incentive effectively (Daumann, 2012). As for the former, the clubs have a credible interest in maintaining a peaceful atmosphere. Otherwise fans, sponsors, and media might choose to avoid the matches. Moreover, clubs can either employ private security companies or reimburse the costs of police operations to effectively prevent violence in and around the stadiums. Therefore, it should be in the clubs' own interest to provide funding for security. Critics might argue that this procedure is incompatible with the constitution, as public safety is a sovereign right and obligation. However, from

[32]These calculations are based on an estimated reduction in assaults per million population per day, 335 days of the football season, and a population of 20,93 million in the affected regions. The number of prevented assaults for the 2014/15 season is: $2.677 \times 335 \times \frac{20{,}930{,}000}{1{,}000{,}000} = 18,770$.

[33]To calculate the annual social cost, I use an estimated cost of 5,067 euros (in 2020 prices) for one assault (Glaubitz *et al.*, 2016).

an economic perspective, it is helpful to distinguish between stadium grounds and public space when discussing the reimbursement of costs for police operations (Mause, 2020). On stadium grounds, a football match can be considered a private good.[34] Accordingly, consumers and producers should pay to ensure security. In contrast, socialization of police costs is justifiable in the public domain as all members of the society benefit equally. This regulation, the assumption of police costs on the club premises, is consistent with the statutes in France, Switzerland, and Great Britain. From the point of view of welfare economics, both positive and negative externalities need to be considered and it must be analyzed whether football matches increase social welfare. However, many positive effects of football matches are not externalities in the strict sense as they have already been internalized.[35] Therefore, an interesting task for future research is to assess all relevant externalities, positive and negative alike, and evaluate whether professional football games generate a positive social net utility.

[34]The attendance of a football match can be regarded as a private good as both conditions of rivalry (each seat can only be sold once) and excludability (the host can deny some people access) are satisfied.

[35]Many economic agents have to pay a fee to be included in the value chain of professional football, such as television companies that pay fees to be allowed to broadcast the sporting events.

B Appendix to Chapter 2

B.1 Figures

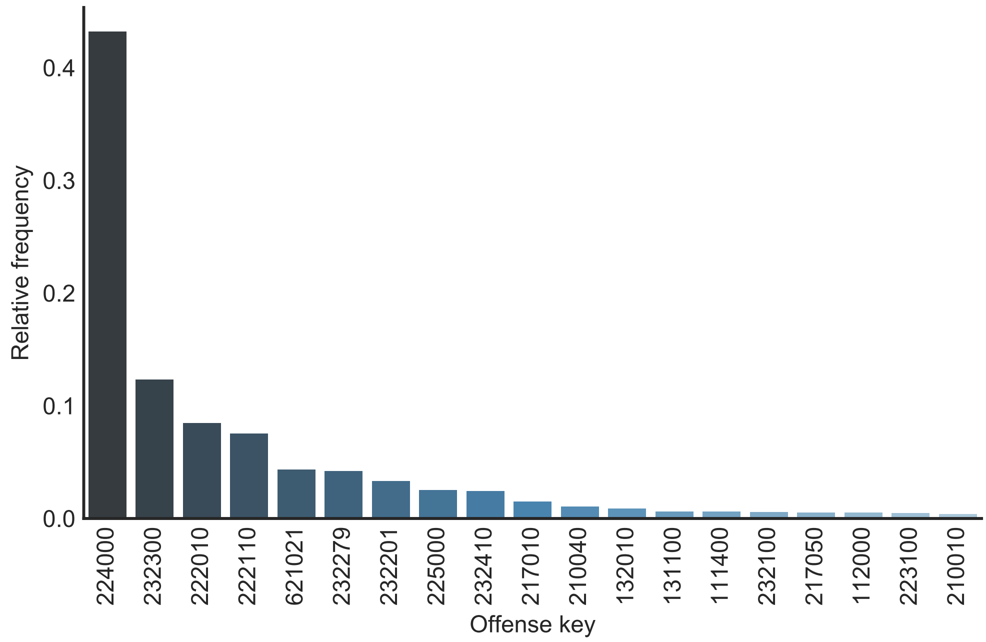


Figure B.1. The types of criminal offenses
Notes: The figure depicts the frequency distribution of the most common criminal offenses in the Federal Republic of Germany in 2014. The most common offense type is simple willful bodily harm (224000), followed by threats (232300), and two forms of dangerous and serious bodily injury (222110 & 222010). These four offense types together comprise around 75% of all criminal offenses.

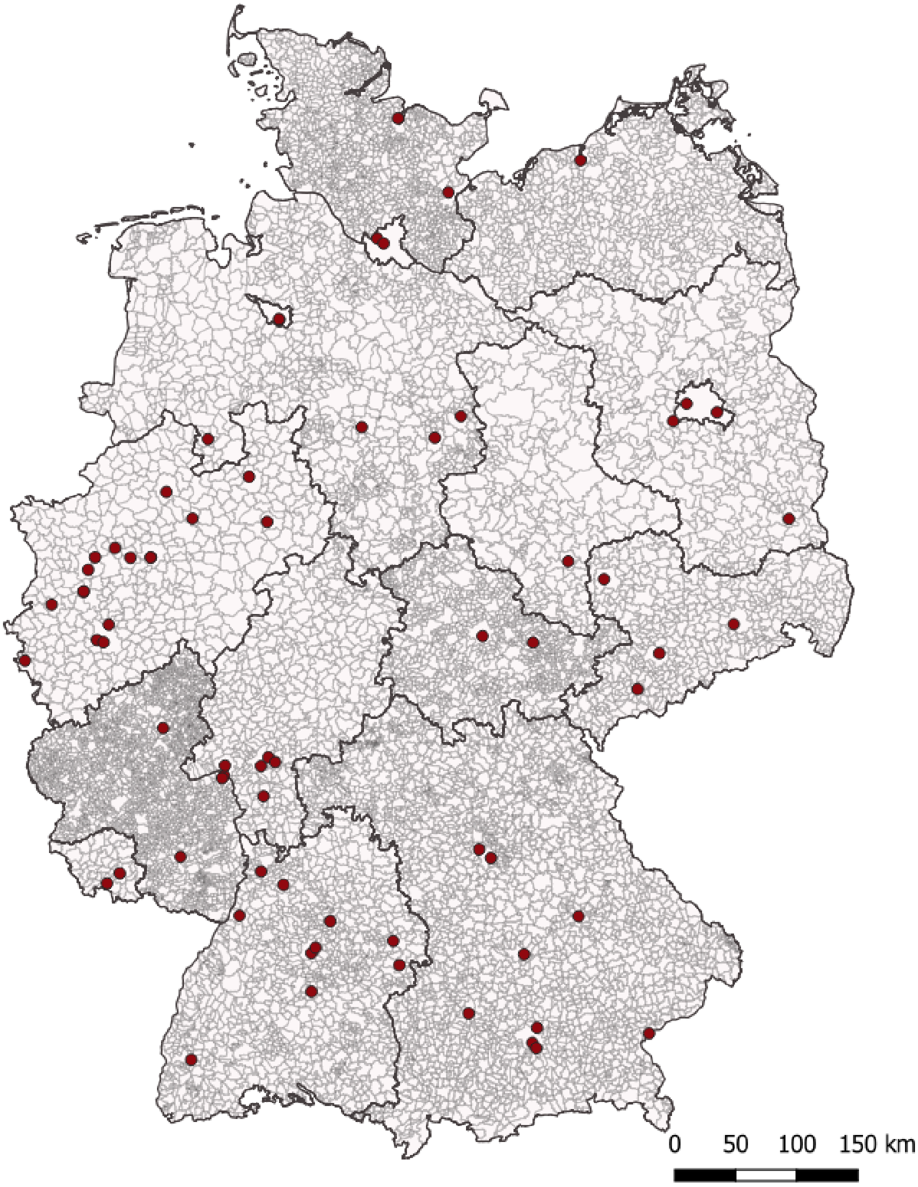


Figure B.2. The stadiums

Notes: This map shows the stadiums used in the analysis over the seasons 2010/11 until 2014/15. The black outlines indicate federal state boundaries.

Source: Own representation with data from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).

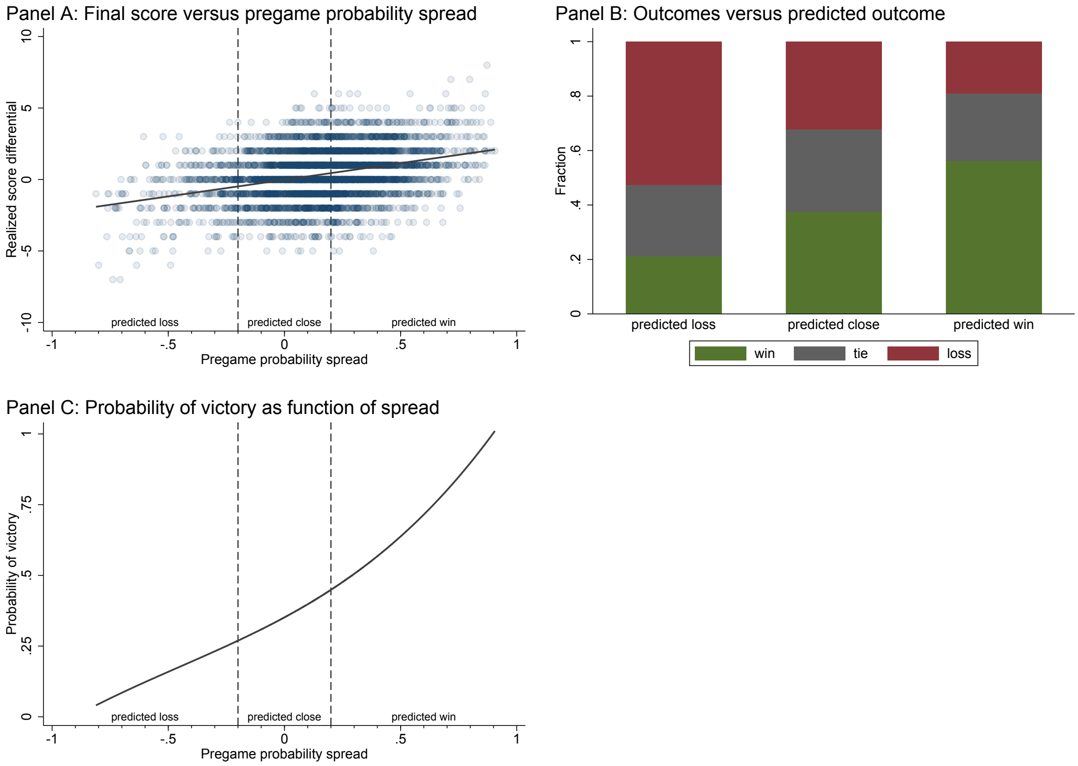


Figure B.3. Pregame probability spread and actual game outcomes

Notes: Panel A shows the relationship between realized score differential versus the pregame probability spread. The realized score differential is defined as the home team's minus the guest team's final score. The plotted regression line has an intercept of -0.020 (s.e. =0.29) and a slope of 2.328 (s.e. = 0.095). Panel B presents the fraction of actual game results by predicted outcome classifications. Panel C shows the probability of winning a game as a function of the probability spread. The curve is obtained from a regression using a third-order polynomial.

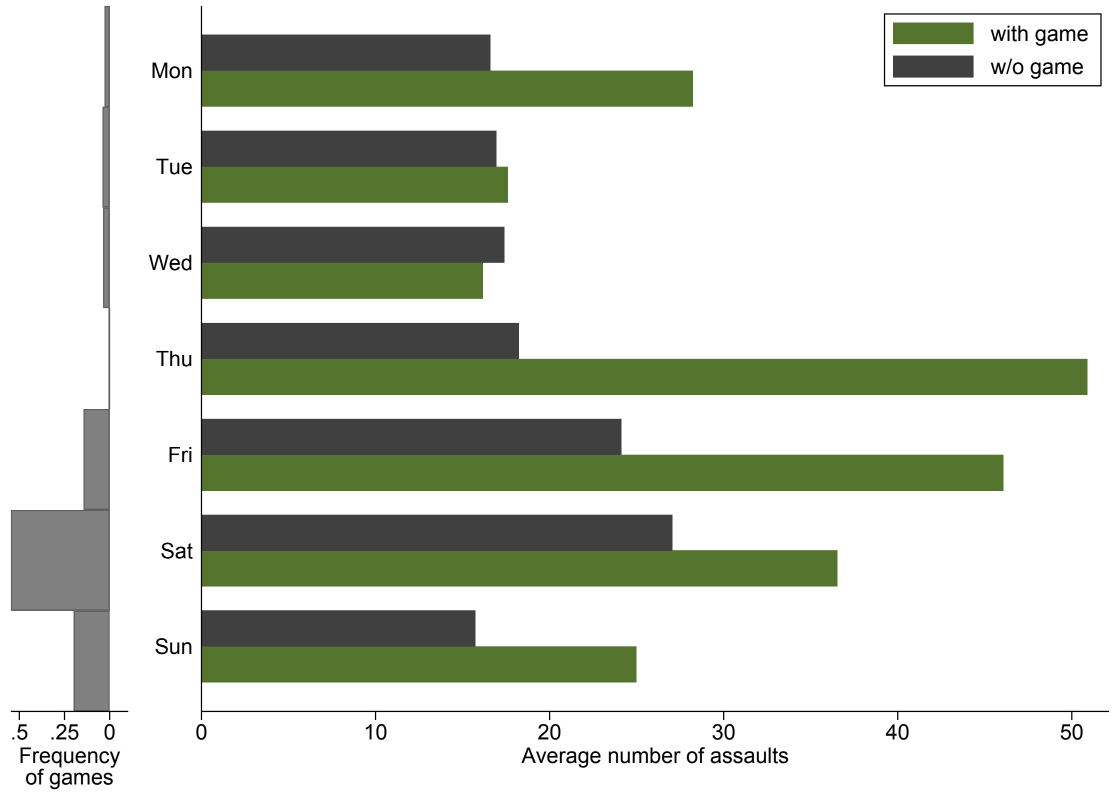


Figure B.4. Average number of assaults on gamedays and days when no game takes place
Notes: The figure shows the daily average number of assaults for regions that host games of a football team from the top three leagues of the German football league system. The daily rates are shown for weeks in which a game is played and for weeks in which no game takes place.

B.2 Tables

Table B.1. Coding of various offenses

	(1) offense key	(2) § StGB
Broadly defined assault		
Simple willful bodily harm	224000	223
Negligent bodily harm	225000	229
Dangerous bodily harm	222010,222110	224
Grievous bodily harm	222020,222120	226
Brawls	222030,222130	231
Resistance to enforcement officers	621021,621029	113
Threatening commission of serious criminal offense	232300	241

Notes: The table shows how the keys of the Police Crime Statistics are translated into the corresponding paragraphs of the German Criminal Code (*StGB*).

Table B.2. High-rivalry matches

(1) team A	(2) team B	(3) comment
Aalen	Heidenheim	Ostalbderby
Aue	Dresden	Sachsenderby
Bielefeld	Münster	Westfalenderby
Braunschweig	Hannover	Niedersachsenderby
Bremen	Hamburg	Nordderby
Dortmund	München	'German Clasico'
Dortmund	Schalke	Revierderby
Dresden	Rostock	Ostderby
Düsseldorf	Köln	Rheinderby
Düsseldorf	Gladbach	Rheinderby
Erfurt	Jena	Thüringenderby
Frankfurt	K'lautern	Südwestderby
Frankfurt	Mainz	Rhein-Main-Derby
Frankfurt	Nürnberg	Derby
Fürth	Nürnberg	Frankenderby
Gladbach	Köln	Rheinderby
Hertha	Union	Berlinderby
Köln	Leverkusen	Rheinderby
Köln	Schalke	Derby
München	Nürnberg	Bayernderby
Münster	Osnabrück	Derby
Rostock	St. Pauli	Derby

Notes: The table shows prominent matches between teams that are known rivals. The above mentioned fixtures make up almost 2.5% of all matches in the sample.
Source: Spiegel (2020), 90min.de (2020), derbys.org (2020)

Chapter 3

The Power of Youth: School Strikes for Climate and Electoral Outcomes*

3.1 Introduction

While the scientific case for mitigating climate change is clear, the political will for bold action remains limited. One pathway out of this dilemma is for voters to no longer accept the risks of climate-related problems and demand political change. Nevertheless, because of the long-term character of climate change risks and the likely delay before mitigating policies will pay off, the support for decisive action against climate change can differ substantially across generations. Indeed, as opposed to older generations, today's young are typically much less likely to accept the status quo as they will experience the brunt of projected mid-century impacts of climate change (Hersch and Viscusi, 2006). Children and youth, in particular, have often been at the forefront of demanding adaptive measures against climate change. Already in 1992, during the UN Climate Conference in Rio de Janeiro, the then 12-year old Severn Cullis-Suzuki addressed delegates by stressing that:

> *"You must change your ways; [...] losing my future is not like losing an election or a few points on the stock market."*

Since then, the consequences of human-caused climate change have become ever more visible, as evidenced by new temperature records almost every year, the melting of the polar ice caps, and the appearance of more extreme weather phenomena. Tellingly, over the course of 2019, Greta Thunberg, the Swedish teen climate activist, has inspired especially young people around the globe to stage some of the largest environmental protests in history. For instance, on Friday, September 20, 2019, roughly four million protesters in more than 150 countries demanded action to fight climate change during the Global Week for Future.

Against this background, our aim is to examine whether school strikes that recently set the tone for environmental protest action worldwide (*Fridays For Future*, henceforth FFF) led to political and economic change. In economics, an extensive literature examines

*This chapter is based on joint work with Helmut Rainer and Maria Waldinger.

the transmission of values and attitudes from adults to children and its consequences for children's behavior (Bisin and Verdier, 2001, Figlio *et al.*, 2019). By contrast, our hypothesis centers around a reverse intergenerational transmission mechanism: youth's participation in environmental protests may raise environmental awareness among their parents' generation and society as a whole. This, in turn, may trigger consumption changes towards more environmentally-friendly products, cause adjustments in mobility and travel behavior, and lead to higher demand for 'green politics' aimed at mitigating the adverse effects of climate change.

The specific question we attempt to answer in this paper is: Did the mobilization of young people to the climate cause have a discernible impact on electoral outcomes in general and on the demand for green politics in particular? We strive to shed light on this question by studying the spread of FFF climate strikes in Germany. Inspired by Greta Thunberg's *Skolstrejk för klimatet* in front of the Swedish parliament in August 2018, FFF quickly turned into a large social movement of school students who skip classes, mostly on Fridays, to participate in mass strikes over climate change inaction. In Germany, the first few climate strikes took place in mid-December 2018, and the FFF movement then gained significant momentum throughout 2019, staging thousands of local climate strikes across the country. The strike activity reached its peak during the Global Week of Climate action in September, when roughly 1.4 million individuals participated in climate protests.

In the first part of the paper, our core contribution is to develop a time-varying, municipality-level measure of youth's propensity to participate in climate strikes. This is a non-trivial challenge due to the geography of mass protests: while strikes are often organized in some central location (e.g., the main city of a region), its participants typically originate both from within and outside that location (e.g., neighboring or more distant municipalities). We tackle this challenge in three steps. First, based on information we collected from police forces, city councils, municipal authorities, and social media, we build up a database consisting of more than 2,300 climate strikes in over 330 German cities during 2019. Second, exploiting cell phone-based tracking data consisting of daily counts of journeys between 260,000 origin-destination pairs, we compute the excess mobility between locations by taking residuals from regressions of the number of journeys between origin i and destination j on day t on a rich set of fixed effects and controls. Third, merging the daily origin-destination mobility residuals with our strike database, we obtain a measure of protest participation that proxies for each strike the geographic distribution of its participants. Crucially, this measure allows us to trace out how protest participation in support of FFF evolved across time and space in Germany. We perform several sensitivity tests to assess the robustness of our strike participation measure and validate our approach for identifying the geographic distribution of individuals attending mass events by examining professional soccer games.

In the second part of the paper, we use time-series and municipality-level variation in our protest participation data to study the role of youth environmental activism in adults' voting behavior. Over the course of 2019, Germany saw four elections: a nation-wide EU election in spring and three federal state elections spread out over the second half of the year. In all four ballots, and especially in the nation-wide EU election, voter turnout increased substantially compared to previous elections. In three out of the four elections, this increase was accompanied by a sharp rise in the vote share of the Green Party. In particular, during its EU election campaign, one of the Green Party's main objectives had been to label the ballot as a *climate election*. On the night of the election day, Michael

Kellner, the party's Secretary-General, announced that he was delighted that his party was able to turn the election into a 'Sunday for Future'. Against this background, we investigate two municipality-level electoral outcomes: voter turnout and the share of votes cast for the Green Party. To purge unmeasured municipality-level, time-invariant factors that may determine a general tendency to vote 'green', we implement a first-differences strategy that compares electoral results in different years (2015 vs. 2019).

Our first set of results shows a positive, large and significant correlation between youth's propensity to participate in climate strikes and voter turnout. In terms of magnitude, a one standard deviation increase in our strike participation index is associated with roughly eight percent of a standard deviation higher voter turnout. Putting this further into perspective, in the context of the elections we study, the estimated correlation corresponds to 3.5 percent of the average increase in voter turnout between 2015 and 2019.

For our second set of results, we establish that in municipalities where youth were highly active in climate protests, the Green Party witnessed larger vote gains compared to previous elections than in municipalities with a lower degree of youth environmental activism. In quantitative terms, we find that a one standard deviation increase in our strike participation index maps into a five percent of a standard deviation higher vote share for the Green Party. For the elections we study, the estimated correlation amounts to roughly 2.5 percent of the Green Party's average vote gains between 2015 and 2019. Taken together, we interpret this evidence as suggesting that youth's protest participation in support of FFF played a non-negligible role in (i) encouraging adult voters to take to the polls and (ii) their demand for green politics.

Our paper touches upon several strands of literature. First, while a substantial part of the literature in economics has highlighted how older generations transmit preferences, norms, and beliefs to younger generations (Bisin and Verdier, 2001, Fernández *et al.*, 2004, Fernández and Fogli, 2009, Figlio *et al.*, 2019), there only exist a few studies, mostly outside of economics, that has looked into the reverse intergenerational transmission. This body of work has established, *inter alia*, that younger generations influence their parents' attitudes towards a variety of controversial topics, including unhealthy consumption behaviors (Flurry and Burns, 2005), the use of modern technology (Baily, 2009), and views on sexual orientation (LaSala, 2000). A particularly relevant subset of studies has explored whether children can foster climate change concerns among their parents. Based on a controlled trial in the Seychelles, Damerell *et al.* (2013) show that adults exhibit more comprehensive knowledge of wetlands and improved water management behavior when their child has received wetland-based environmental education. In a similar vein in the United States, Lawson *et al.* (2019) present an experimental evaluation of an educational intervention program designed to build climate change concern among parents through their middle school-aged children, finding that parents of children in the treatment group expressed higher levels of climate change concern than parents in the control group.

Second, our paper has antecedents in a small but insightful literature that examines the impact of social and political movements. Studying the Tea Party protests in the United States, Madestam *et al.* (2013) establish that the movement has caused a shift to the right in policymaking, both directly through incumbent politicians' decision-making and indirectly through voters' selection of politicians in elections. Using daily variation in the number of protesters during Egypt's Arab Spring, Acemoglu *et al.* (2018) show that more intense protests are associated with lower stock market valuations of firms connected to politicians in power relative to unconnected firms. In addition, several papers have studied the determinants of protest participation (Finkel and Opp, 1991, Finkel and Muller, 1998, Cantoni *et al.*, 2019, Bursztyn *et al.*, forthcoming). An interesting but not directly related phenomenon is described by Depetris-Chauvin *et al.* (2020): collectively shared experiences of the type induced by mass sports events—in particular, international football games in the context of Sub-Saharan Africa—can shape identities in ways that can help build national sentiment at the expense of ethnic identification.

Finally, our work relates to studies that use cell phone data to examine economically and socially important phenomena. A large part of this literature centers around the investigation of social network patterns (Onnela *et al.*, 2007, 2011, Kovanen *et al.*, 2013). Beyond this, cell phone records have recently been used to predict the spatial distribution of urban economic activity from commuting choices (Kreindler and Miyauchi, 2021) and to assess the contagion externality of mass events (Dave *et al.*, 2020). We instead exploit cell phone-based tracking data to measure the spread of a major social movement across time and space.

The remainder of the paper proceeds as follows. In Section 2, we provide background information on the Fridays for Future movement. Section 3 contains a description of the data we use. In Section 4, we lay out our methodology for measuring youth's climate protest participation across time and space. In Section 5, we explore the role of youth environmental activism in adults' voting behavior, discussing our empirical strategy before presenting the results. The final section concludes.

3.2 Background

Fridays for Future is a global social movement consisting to a large extent of school students that claim comprehensive, fast, and efficient measures for climate protection. The students skip school, mostly on Fridays, to lay emphasis on their key demand: to adhere to the targets set in the 2015 Paris Agreement to lower global carbon dioxide (CO_2) emissions in order to limit the associated increase in global warming to 1.5 degrees above pre-industrial levels. Parts of the scientific community (*Scientists for Future*) have stood in solidarity with FFF, declaring the concerns as justified and expressing support for the demands put forward by FFF (Warren, 2019, Hagedorn *et al.*, 2019). The movement is organized in local chapters that plan weekly school strikes for the climate and regular large-scale coordinated marches (Smith and Bognar, 2019). In 2019, there have been four large-scale global climate strikes in March, May, September, and November. For instance, alone for the third global climate strike in September 2019, FFF organized 6,000 protests in 185 countries, which mobilized 7.6 million participants (De Moor *et al.*, 2020). The exact dates of the global climate strikes and other relevant events are presented in the timeline in Panel (a) of Figure 3.1.

The Role of Greta Thunberg.—The movement received its initial impulse from Greta Thunberg, the then 15-year old Swedish teenager, who started her *Skolstrejk för klimatet* (school strike for climate) in front of the Swedish parliament in August 2018. Her actions quickly received media attention around the world. Greta Thunberg soon started to advocate her goals at important international events such as the 2018 Climate Change Conference in Katowice, the World Economic Forum in Davos in January 2019, in front of the European Parliament on April 16, or the United Nations (UN) climate change con-

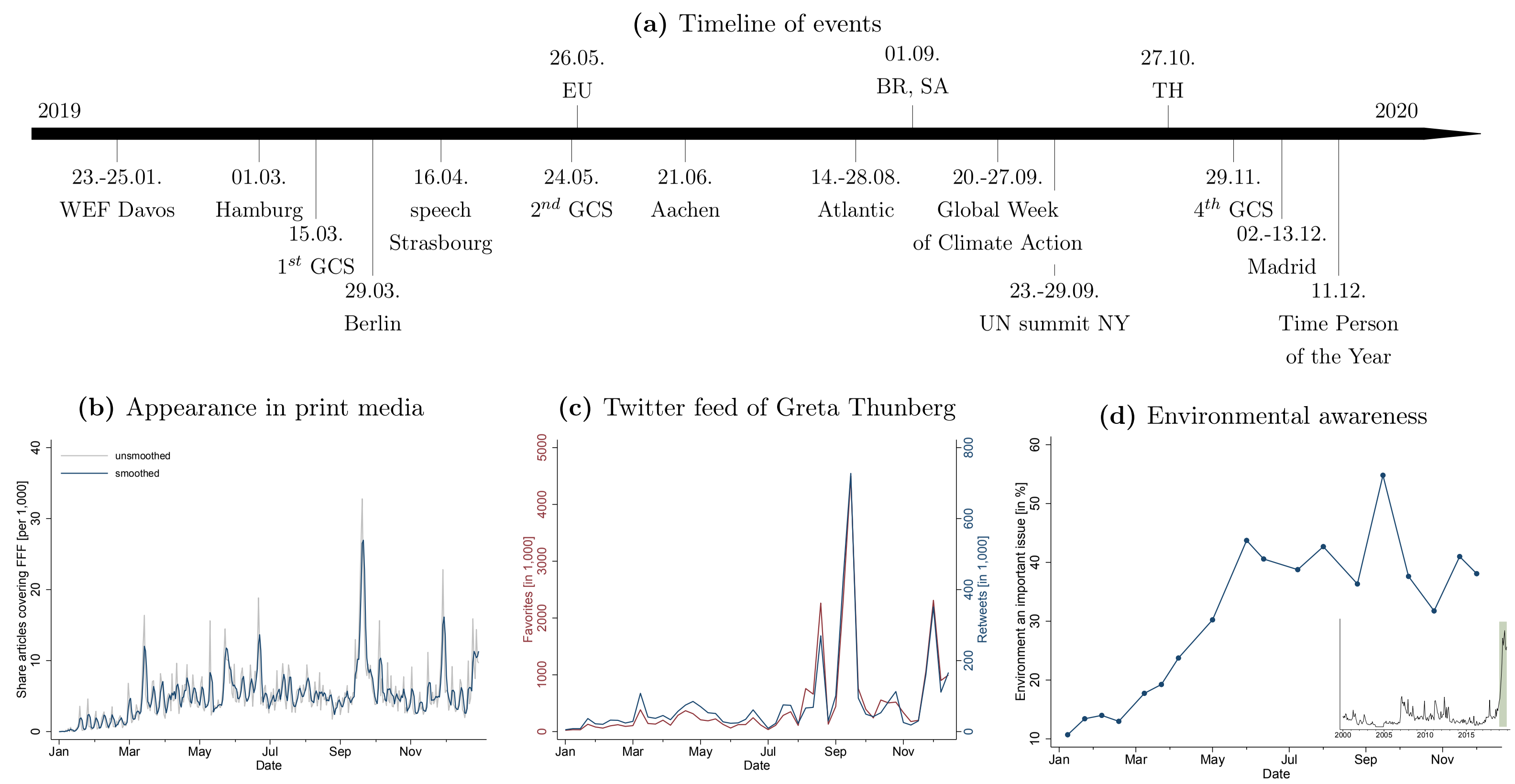


Figure 3.1. Important events surrounding FFF, perception in (social) media, and environmental awareness in 2019

Notes: Panel (a) presents a timeline of relevant events over the year 2019. Above the timeline, the scheme lists elections that fall in the sample period (election for the European Parliament, Brandenburg, Saxony, and Thuringia). Below the timeline, there is a series of important events, either far-reaching strikes or critical milestones for Greta Thunberg. Panel (b) shows the daily share of print articles covering FFF (per thousand articles). An article is defined to report about FFF if it contains at least the keyword 'Fridays for Future' (in various notations) or 'climate strike'. The blue line stems from a moving average with a window of three days. Panel (c) plots the weekly number of favorites and retweets of Greta Thunberg's tweets (in thousand). Panel (d) presents the share of respondents who state that the environment is the most important issue in the representative election survey Politbarometer (2019). Since 1977, the survey informs about the political mood, projections, and attitudes towards current topics.
Source: Own representation with data from Genios, Twitter, and Forschungsgruppe Wahlen.

ference in Madrid in December 2019. She sailed across the Atlantic to deliver her famous speech at the UN Climate Action Summit in New York in September 2019 ('how dare you?'). The movement's success is tightly related to her, not only on the world stage but also on a more local, personal level. A 'Greta Effect' has been documented in many dimensions: it ranges from increases in carbon offsetting projects (The Guardian, 2019b) to a reduction in the number of airline passengers due to 'flight-shaming' (The Economist, 2019). Most importantly, Greta Thunberg acted as a role model for many young people's decision to take part in the climate strikes. During the first global climate strike on March 15, roughly 45 percent of German students stated that Greta Thunberg influenced their decision to participate in the event (Sommer *et al.*, 2019). The effect of Greta Thunberg as a role model has been more pronounced for girls than for boys. Greta Thunberg's commitment to fighting for climate change mitigation was acknowledged by various prestigious honors and awards such as the Right Livelihood Award (promoted as the 'Alternative Nobel Prize'), being the youngest 'Time Person of the Year' in 2019, and two consecutive nominations for the Nobel Peace Prize (in 2019 and 2020).

The German Context.—In Germany, the first strikes started in mid-December 2018 in a handful of cities.[1] By mid-January 2019, there were already 25,000 protesters in 50 places, according to the organizers of FFF. The movement gained momentum in March due to Greta Thunberg's visits in Hamburg and Berlin on March 1 and 29,[2] and the first global climate strike on March 15. On that day, there were an estimated 300,000 protesters on German streets, and almost 1.8 million protesters participated worldwide. The second global climate strike took place just before the 2019 elections for the European Parliament in order to allow a significant impact on electoral outcomes (Smith and Bognar, 2019). FFF had declared the 2019 EU election as the 'climate election' (#voteforclimate) and green parties obtained historically good results all across Europe. In Germany, there were 218 events (more than in any other country) and different election surveys indicated that environmental issues and the topic of climate change affected voters in the election (Time, 2019). The first international climate strike took place in Aachen on June 21. The strike activity culminated in the Global Week of Climate Action in September. During that week on September 20, the day of the third global climate strike, the largest climate strikes in world history were reported (The Guardian, 2019a). While more than 7.6 million individuals participated in climate strikes globally, the numbers amounted to 1.4 million protesters in 550 cities in Germany (alone in Berlin there were almost 300,000) (De Moor *et al.*, 2020). The strikes were timed to increase pressure to take action mitigating climate change ahead of the UN Climate Action Summit in New York. On September 20, the German government also announced a 54 billion climate change package, which included the introduction of a price per ton of emitted CO_2. The fourth and last global climate strike in 2019 happened on November 29, again just days before the UN Climate Change Conference started in Madrid. The estimated strike participation had declined and about 630,000 individuals joined the strike in Germany (Zeit Online, 2019).

Who Are the Participants?—The results from surveys conducted at strikes allow drawing a profile of FFF participants (Sommer *et al.*, 2019, De Moor *et al.*, 2020). During the first global climate strike in March 2019, almost three out of four protesters were students (school or university), and 60 percent of participants were female. Roughly 70 percent of

[1] The following information on the German context stems, except where otherwise noted, from Sommer *et al.* (2019).

[2] The two strikes that were vistited by Greta Thunberg in March 2019 attracted large populations. According to the organizers, there were 10,000 students in Hamburg and 25,000 people in Berlin.

the respondents located themselves at the center-left of the political spectrum which is far above the value for the general population. Many participants come from upper-middle-class households, as indicated by high levels of (parental) educational attainment. The share of mothers (fathers) with a university degree was 45.8 (49.4) percent, values that are twice as high compared to the population average. The protesters are on average largely inexperienced with demonstrations (40% no experience, 42% attended few demonstrations) and were mobilized through either social media (37%) or interpersonal channels (31.5%), such as friends and schoolmates.

The Importance of (Social) Media.—The media coverage on the topic of FFF was crucial for the movement's success. The first coverage in Germany started already in the late autumn of 2018. While the initial coverage was centered around the violation of compulsory school attendance, it soon reported more on the contents of the demands made by FFF. Panel (b) of Figure 3.1 shows the share of articles covering FFF in German print media over the year 2019. The data is obtained via webscraping from Genios that provides an archive containing almost 300 daily and weekly outlets in Germany. There are clear spikes in media coverage around each of the global climate strikes in March, May, September, and November. The maximum share of articles is attained during the Global Week of Climate Action: more than 3.2 percent of all print articles in Germany were reporting on FFF. The same tendency holds for social media. Panel (c) of Figure 3.1 shows the weekly number of favorites and retweets of Greta Thunberg's tweets. Again, the attention in Greta Thunberg's tweets culminates with the global climate strikes. The Appendix gives further insights into the content of Greta Thunberg's tweets.[3]

Public Opinion.—Related to the increase in media coverage is a rapid shift in public opinion on environmental issues (Smith and Bognar, 2019). The Institute for election research (*Forschungsgruppe Wahlen e.V.*) has been conducting the survey Politbarometer (2019) for more than 40 years, which regularly polls attitudes with regard to political parties, current issues, and delivers projections for upcoming elections. One question in the survey asks about respondent's opinion on the two most pressing current issues in the Federal Republic of Germany. The large graph in Panel C of Figure 3.1 shows that the share of respondents who report that the environment as a high-priority issue is increasing significantly from around 10 percent to almost 60 percent over the year 2019. The small graph plots the long-run development since 2000 (the year 2019 is highlighted in green). Before the year 2019, an average share of four percent reported that the environment is among the top two pressing issues.[4] Respondents also reported that they expect the climate strikes and the associated sudden spike in environmental awareness would lead to political changes (Forschungsgruppe Wahlen e.V., 2019). While 37 percent of respondents expected an impact on politics in April 2019, the share was already 51 percent at the end of June.

[3]Figure C.1 gives an overview of the most widely used words, hashtags, and mentions. Figure C.2 displays information on the sentiment and the length of the tweets. Figure C.3 plots the weekly number of favorites of influential German climate activists, such as Luisa Neubauer.

[4]Before 2019, a maximum share of 14.5 percent of respondents reported the environment as a top issue.

3.3 Data

Our main sample uses data from 2019 and contains all German regions, either on the county (*Kreise*, $N = 401$) or municipality (*Gemeinde*, $N = 10,719$) level. We combine six sources of data to residualize journeys and to investigate associations between climate strike participation and electoral outcomes.

3.3.1 Cell Phone-Based Tracking Data

The cell phone-based tracking data is obtained from *Teralytics*, which draws on the universe of customers from the *Telefonica* O_2-network. The mobile network provider had a market share of 31 percent in 2019 (Statista, 2020) and is a popular carrier among young people.[5] *Teralytics* uses machine learning-based technology to transform mobile signals into a number of journeys between two locations.[6] The data reports the daily number of trips between origin-destination pairs, which is the unit of observation in the first part of the empirical analysis. Over the year 2019, the data contains 64.4 billion trips in total. Due to data privacy regulations, *Teralytics* cannot report journeys if there are fewer than five trips for a given origin-destination pair per day, resulting in missing observations. The spatial level of aggregation mostly coincides with county borders. Only for larger metropolitan areas, there are regions with smaller clusters. For this reason, the data set provides us the number of journeys for 513 different geographies, which result in more than 260,000 possible origin-destination pairs.[7] The vast majority of trips (92.7 percent) do not exceed 30 kilometers.[8] For larger distances, the data can distinguish between different modes of transportation.

3.3.2 Climate Strike Data

Our data on climate strikes is self-collected and comes from local authorities, social media, and the website of FFF Germany. All public gatherings such as rallies and demonstrations must be registered in Germany. The registration usually happens at the police, the city council, or other regulatory agencies. The large bulk of our database is retrieved by contacting the respective authorities and requesting a list of all climate strikes that took place in their jurisdictions in 2019. Consequently, we obtained information on the exact location and time of nearly 2,000 strikes. In the next step, we used social media postings (Twitter, Facebook, and Instagram) to fill up missing information for the 20 largest cities that were missing in the database. Searching for information on social media channels gave us almost another 400 climate strikes and ensured that we do not miss any large-scale climate strike. In the last step, we enriched our database with information on almost 1,600 strikes retrieved from the website of FFF Germany. Using the internet archive of `www.web.archive.org`, we could restore snapshots with information on climate strikes for

[5] To obtain representative mobility patterns, *Teralytics* extrapolates to the entire population with the help of regional market shares.

[6] There is a required minimum time between individual movements so that they count as a completed journey, depending on the mode of transport: It is 30 minutes by car, 60 minutes by train, and 120 minutes by plane.

[7] For 355 counties, the geographies overlap with the county borders. The remaining 46 counties are split into subclusters, with a maximum of five subclusters for the largest metropolitan areas.

[8] The distance refers to the linear distance between the centroids of two geographies.

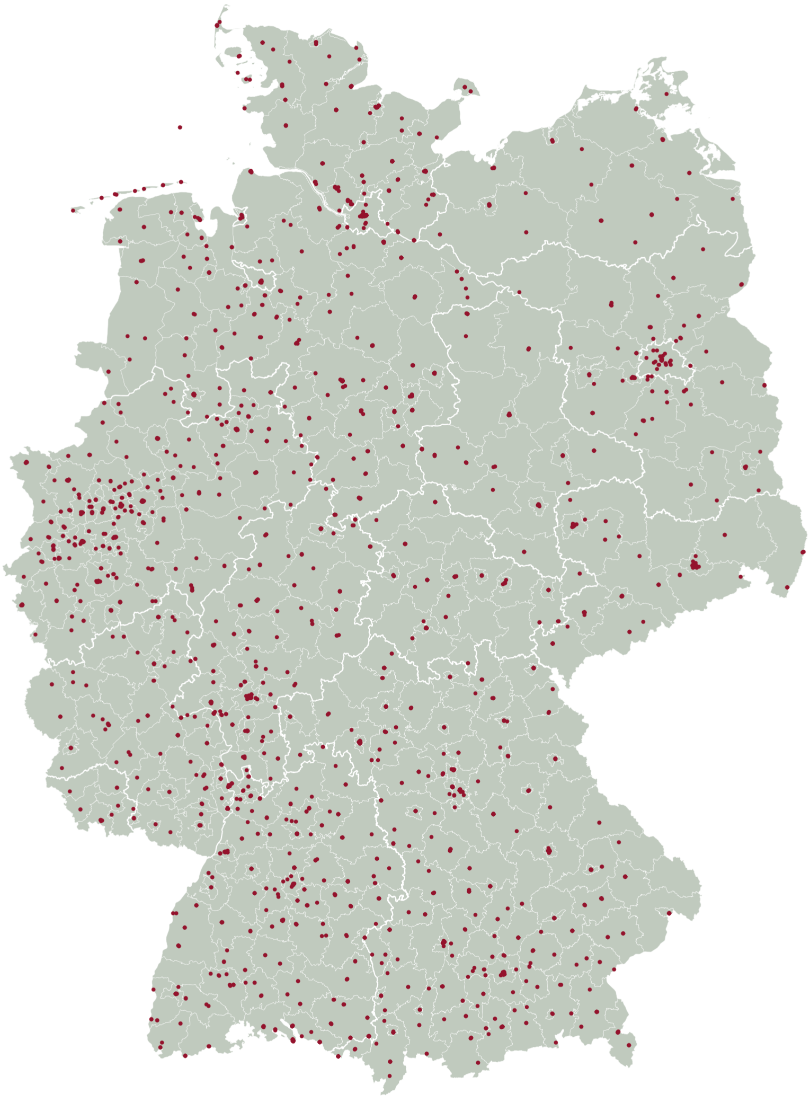


Figure 3.2. Locations of climate strikes in 2019
Notes: The map shows the climate strikes (red dots) in our data base over the year 2019. The bold white lines indicate state boundaries and the thin white lines represent county boundaries.

some weeks. After combining the information from the different sources, the locations where the climate strikes took place were geographically encoded. Figure 3.2 shows a map with all strikes contained in our database. There is considerable variation in the strike behavior across the year as depicted in the maps in Appendix Figure C.4. The same impression is obtained from Figure 3.3, which presents the daily number of strikes across the three sources. The highest number of strikes is reported on the global climate strike events (March, May, September, and November). Furthermore, it is visible that the strikes

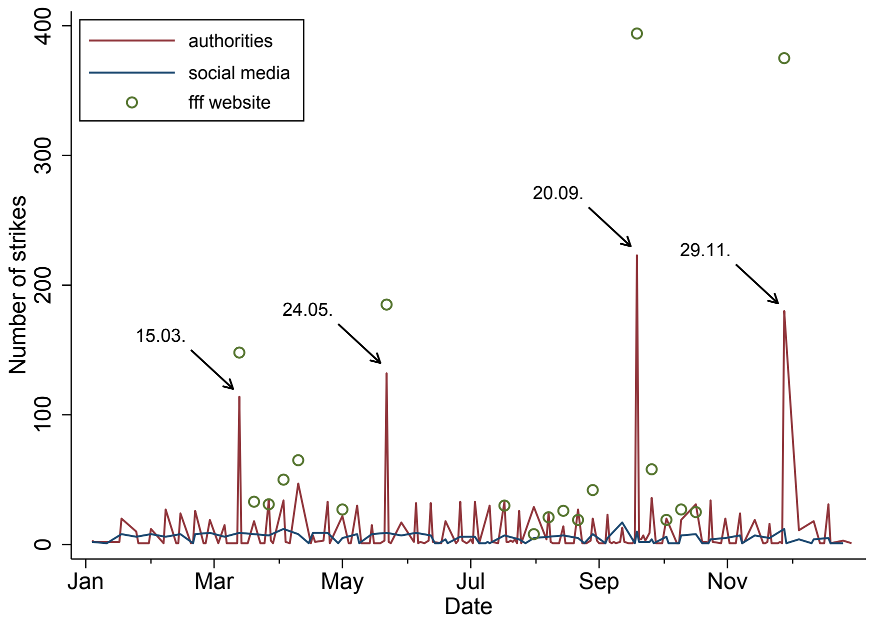


Figure 3.3. Number of strikes across sources
Notes: The figure shows the daily number of strikes per source. The indicated dates above the spikes mark the four global climate strikes.

from the FFF website are only available in 19 weeks when the domain was archived. The other sources, however, exhibit more continuity in how they report the number of strikes.[9]

3.3.3 Electoral Data

For the European Parliament election, we use data provided by the Federal Statistical Office and the statistical offices of the Länder. For the state elections in Brandenburg, Saxony, and Thuringia, we draw on data from the State Returning Officers (*Landeswahlleiter*) and the statistical offices of the Länder. In our analysis, we use the municipality-level voter turnout and the vote share for the Greens (*Bündnis 90/Die Grünen*) as our dependent variable.[10] The vote share is defined as the number of votes for the Greens divided by the total number of valid votes cast. In addition to the vote shares at the four elections in 2019, we also look at first-differences that use the corresponding counterparts in 2015. In our sample, the average vote share for the Greens is 19.68 percent. Table 3.1 shows that there is substantial heterogeneity across elections.

3.3.4 Weather Data

The weather data is derived from Germany's National Meteorological Service (*Deutscher Wetterdienst*). To construct the weather controls, we aggregate weather information to the

[9] Because of the extreme volatility of the strikes from the FFF Germany website - no information in most instances vs. detailed regional resolution on others-, we first focus on strikes for which there is more continuous information, i.e. strikes obtained from the authorities and social media.

[10] Our analysis excludes municipalities (around 5 percent) that were subject to territorial reforms between the elections 2015-2019.

Table 3.1. Vote share for the Green Party and voter turnout across elections

		Greens (2019)		Δ Greens (2019-2015)		Turnout (2019)		Δ Turnout (2019-2015)	
Election	Obs.	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
EU	10,719	20.76	7.54	9.73	3.86	59.18	7.31	13.06	5.81
Brandenburg	413	10.17	5.12	4.31	2.19	56.31	8.81	9.72	6.42
Saxony	414	4.94	1.91	1.31	0.91	65.32	6.83	16.90	4.22
Thuringia	645	5.32	3.41	-0.57	0.97	61.11	6.62	11.85	3.82
Total	12,191	19.68	8.23	9.11	4.33	59.33	7.42	13.05	5.82

Notes: The Table reports the number of municipalities per election (column 2) and population-weighted means and standard deviations for the Green's vote share in 2019 (columns 3 and 4) and the first-differenced vote share for the Greens (columns 5 and 6). Columns 7 through 10 report the corresponding values for voter turnout.

county level using the weighted average (inverse distances) of the values from the individual weather monitors within a certain radius to the county's centroid.[11] We use daily averages of the following weather variables: hours of sunshine, precipitation, and maximum air temperature.

3.3.5 Holidays

To account for differences in the number of journeys between origin-destination pairs on regular and special days, we include state-level controls for public and school holidays. The data on school holidays is derived from 'The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany' (*Kultusminister Konferenz*). The data on public holidays is collected from www.schulferien.org. Additionally, we include dummy variables for peculiar days (New Year's Eve and Carnival season), which shift the expected number of journeys between origin-destination pairs.

3.3.6 Other Regional Variables

The Federal Statistical Office and the statistical offices of the Länder provide a database (*Regionaldatenbank*) of detailed statistics by different subject areas at very granular spatial levels. We draw on this database for various purposes. First, we extract municipality-level population figures that are used to construct the strike participation index and to weigh the observations when analyzing the associations between strike participation and electoral results. Second, we collect municipality-level information on topics such as per capita income, unemployment rates, and demographic characteristics that are used as controls.

[11] The maximum radius depends on the weather variable. It is 30 km for precipitation, while it is 50 km for hours of sunshine and maximum air temperature.

3.4 Granular Measurement of Strike Participation

3.4.1 Residualize Journeys and Combine with Strike Data Base

In the first part of the empirical analysis, we exploit the cell phone-based tracking data to measure local strike participation. To determine how many participants attend a climate strike and where they come from, we residualize the number of movements between origins and destinations. We exploit a rich fixed effects model to remove the predictable variation in the number of journeys based on the origin-destination pair, date characteristics, holiday fixed effects, and weather controls. Let journeys_{ijt} denote the number of journeys between origin i and destination j at time t, which in our baseline model is given by

$$\text{journeys}_{ijt} = \alpha + \vartheta_{ij} + \underbrace{\varphi_d + \eta_w + \psi_m}_{\gamma_t} + \lambda \mathbf{X}_{ijt} + \varepsilon_{ijt}. \tag{3.1}$$

Origin-destination fixed effects are captured by ϑ_{ij} and account for time-invariant mobility patterns between origin-destination pairs (mostly on the county level with few subclusters). They ensure we use within instead of between origin-destination pair variation over time. The time fixed effects vector γ_t contains binary variables for the day-of-the-week (φ_d), the week-of-the-year (η_w), and the month (ψ_m). It allows to flexibly control for weekly and seasonal mobility patterns. The vector $\mathbf{X}_{ijt}$ includes public as well as school holidays and weather controls, both at the origin and destination geography.[12] The model explains a high proportion of the variance in the number of origin-destination journeys as the adjusted $R^2 = 0.9696$.

To obtain the number and origin of people participating in climate strikes, we calculate the difference between the observed number of journeys and the predicted number of journeys conditional on the variables described above $e_{ijt} = (\text{journeys}_{ijt} - \widehat{\text{journeys}}_{ijt})$. The residuals e_{ijt} capture the excess mobility net of the variation stemming from the baseline covariates. We match the daily origin-destination-level residuals with our strike database in the next step. The climate strike database contains the precise time and location for the set of destinations $j \in \{1, ..., J\}$ in which a strike takes place on $t \in \{1, ..., T\}$ (mostly Fridays). Consequently, e_{ijt} indicates how many people are coming from origin i to a strike in destination j happening on day t.

Figure 3.4 shows intuitively how the resulting strike participation measure looks like for two exemplary strikes in Hamburg and Berlin. Both of the climate strikes were visited by Greta Thunberg and attracted large populations. A darker shade of green indicates that more people in that geography were coming to one of the two strikes (locations marked in red) net of the variation stemming from the controls discussed above. The color scale classification is obtained by applying the Fisher-Jenks natural breaks algorithm. The algorithm searches for natural breaks in the data by minimizing the variance within clusters and maximizing the variance between clusters (Jenks, 1967). In both instances, strike participation is a relatively local phenomenon. The highest excess population flows are coming from adjacent geographies.

[12] State-level public holiday controls contain dummy variables for All Saints' Day, Ascension Day, Assumption Day, Christmas, Corpus Christi, Epiphany, Easter, German Unity Day, Good Friday, Labor Day, New Year's Day, Penance Day, Pentecost, and Reformation Day. Furthermore, it contains binary variables for Carnival season and News Year's Eve. County-level weather controls include air temperature, hours of sunshine, and precipitation.

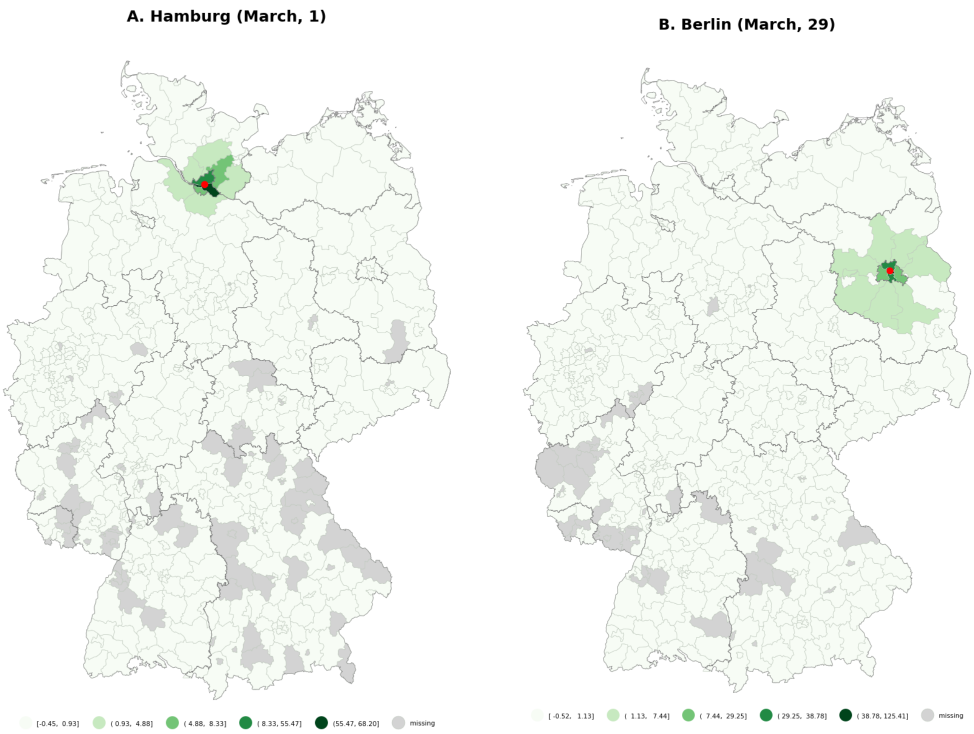


Figure 3.4. Strike participation for selected strikes

Notes: The maps show residualized movements (in thousand) for two exemplary strikes. The two strikes were both visited by Greta Thunberg and attracted large populations. A darker shade of green indicates that more people were coming to the climate strike conditional on the controls discussed in the text. The color scale classification is obtained by using the Fisher-Jenks natural breaks algorithm. The red dots mark the strikes' location, gray areas indicate missing data (censored), bold gray lines show state boundaries, and thin gray lines represent the regions defined by *Teralytics*.

We perform several sensitivity tests to assess the robustness of our local strike participation measure. Figure 3.5 shows maps for the climate strike in Hamburg using alternative participation measures.[13] First, we expand the baseline specification by including interactions of $\vartheta_{ij} \times \gamma_t$ in Equation 3.1, i.e. origin-destination-by-day-of-the-week fixed effects, origin-destination-by-week-of-the-year fixed effects, and origin-destination-by-month fixed effects. The interactions have the advantage of allowing for systematic changes in the origin-destination-level mobility patterns over the year. The top row maps illustrate different versions of included interactions when estimating the equation with OLS. Panel A displays the baseline specification as already shown in Figure 3.4. Panel B adds interactions of the origin-destination fixed effects with week and month fixed effects. We refrain from also interacting the origin-destination fixed effects with the day-of-the-week fixed effects because the interaction would likely absorb the average strike participation in regions where climate strikes happen every Friday. For completeness, we show a fully interacted version in Panel C. Second, we test the robustness of the strike participation measure by estimating Equation 3.1 with a Poisson model to account for the discrete nature of the dependent variable. We do so by also showing the different interaction specifications discussed in the OLS case. Throughout all specifications emerges a picture of remarkable consistency. The models identify roughly the same set of regions where strike participants come from, irrespective of the modeling choices. The levels might be different, but this will not be an issue as we use standardized versions of the strike participation measure.

[13] The corresponding maps for the climate strike in Berlin can be found in Appendix Figure C.5.

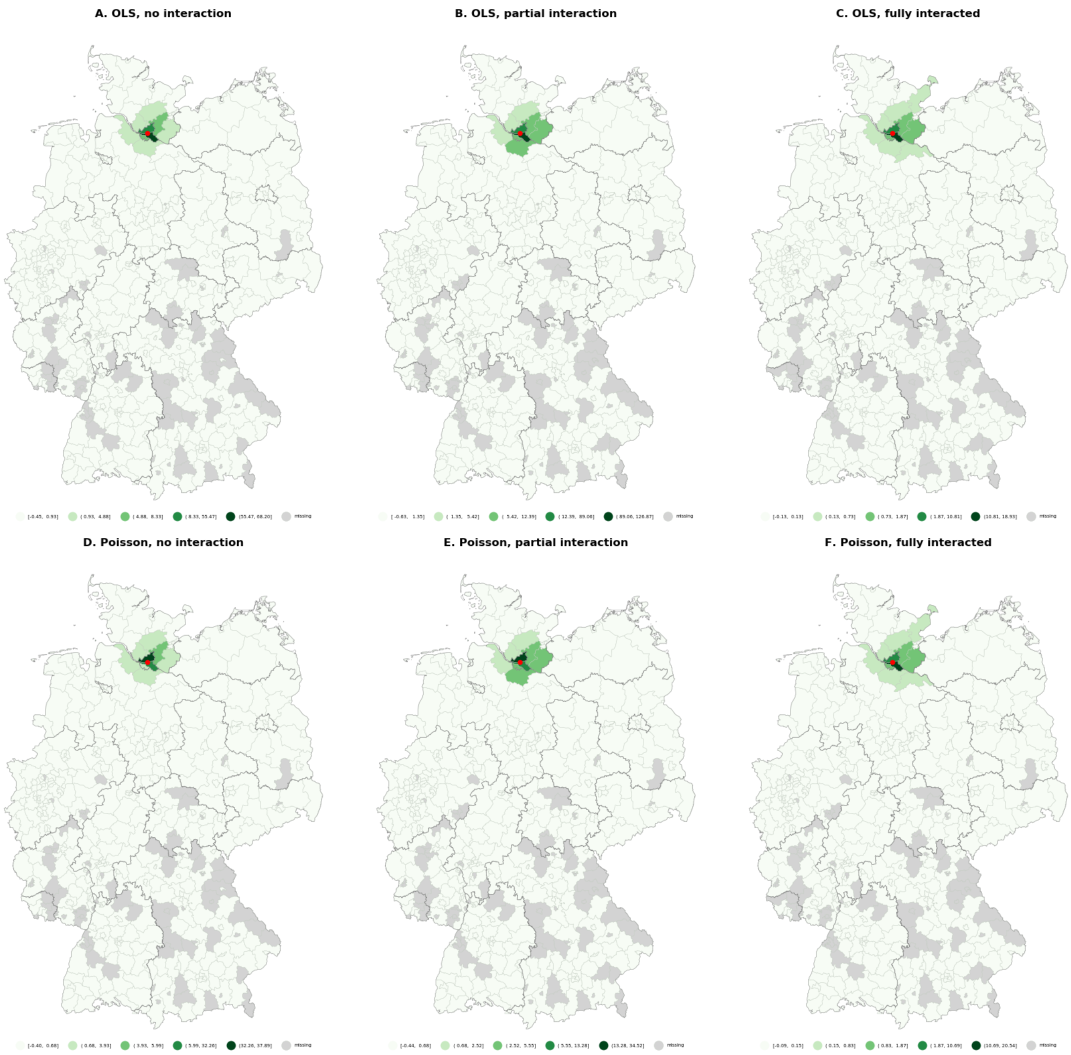


Figure 3.5. Alternative participation measures for the climate strike in Hamburg
Notes: The maps show residualized movements (in thousand) for the climate strike in Hamburg (March, 1). Column 1 uses the baseline specification in Equation 3.1, column 2 further adds $\vartheta_{ij} \times \eta_w$ and $\vartheta_{ij} \times \psi_m$, and column 3 presents a fully interacted version (plus also including $\vartheta_{ij} \times \varphi_d$). The top row uses OLS to estimate Equation 3.1, while the bottom row shows results when using Poisson. A darker shade of green indicates that more people were coming to the climate strike conditional on the controls discussed in the text. The color scale classification is obtained by using the Fisher-Jenks natural breaks algorithm. See Figure 3.4 for additional details.

Throughout the paper, we focus on the baseline specification but also present results for the alternative strike participation measures discussed above.

As a further check, we validate our approach to identify the origin of people attending other large-scale public events by investigating the catchment areas for soccer games. In that respect, Figure C.6 presents results for matches happening in Munich, Dortmund, and Freiburg (an example of a large, medium, and small city). Reassuringly, the maps look similar to those depicting climate strikes and illustrate that the catchment areas lie around the matches' location. However, in comparison to the climate strikes, the catchment areas are a bit more extensive and contain the guest teams' origin.

3.4.2 Construction of a Municipality-Level Participation Index

After identifying the origin of strike participants, we want to account for the intensity of local participation rates over time. The idea is to differentiate between regions with continuous and sporadic strike involvement. In this section, we incorporate these considerations by constructing a cumulative measure of strike participation at the municipality-level. We define our index as follows:

$$\text{Participation Index}_{m\tilde{t}} = \left(\frac{w_{mi}}{\text{population}_m} \sum_{t=1}^{\tilde{t}} \sum_{j=1}^{J} e_{ijt} \right), \tag{3.2}$$

where Participation Index$_{m\tilde{t}}$ is a cumulative measure of relative strike participation in origin municipality m up to a specific point in time $\tilde{t}$ (for instance when an election takes place). It comprises the following elements: $\sum_{t=1}^{\tilde{t}} \sum_{j=1}^{J} e_{ijt}$ indicates the cumulative strike participation in county i up to time $\tilde{t}$.[14] We sum over destinations j to account for the fact that individuals residing in i are going to climate strikes in more than one destination. Furthermore, we sum over all days up to and including $\tilde{t}$ in which county i appears as an origin for a strike in any destination j. The weights w_{mi} allocate the strike participation from the county to the municipality level. They are defined as the share of children (aged 0-17) living in municipality m in relation to other municipalities in county i. The weights reflect the notion that strike participation should be higher in municipalities with many young people relative to municipalities with few young people. We restrain our weights to minors, i.e. the population share that is not allowed to vote, with the purpose to not generate associations with electoral outcomes by construction. Finally, we take the number of cumulative strike participants in municipality m at time $\tilde{t}$ and divide it by the municipality's population figure to obtain a relative measure.

3.5 Strike Participation and Electoral Outcomes

3.5.1 Empirical Strategy

Our approach to investigating the relationship between strike participation and electoral results is based on a fixed-effects model, which exploits within-state and within-election variation of local strike participation rates. We estimate the following model:[15]

$$\text{Share Greens}_{me} = \theta_s + \tau_e + \beta \cdot \text{Participation Index}_{me} + \mu \mathbf{X}_{1m} + \xi_{me}, \tag{3.3}$$

where Share Greens$_{me}$ is the Greens' vote share in municipality m for election e, which we define as the number of votes for the Greens relative to the total number of valid votes cast. We regress the dependent variable on state and election fixed effects (θ_s and τ_e). The Participation Index$_{me}$ captures the cumulative relative strike participation up to the time when election e takes place. For some specifications, we also include a set of municipality-level controls $\mathbf{X}_{1m}$, such as income, unemployment, and demographic characteristics.

[14] We only select non-negative residuals to obtain positive values of strike participation.

[15] For many aspects in this section's empirical approach, we found inspiration in the study of Cantoni *et al.* (2020) who investigate the municipality-level relationship between the support for Germany's Nazi party in 1933 and the likelihood to vote for the 'Alternative for Germany', an increasingly right-wing populist party.

To eliminate unobserved time-invariant municipality specific effects, such as a general tendency to vote for the Greens, we employ a first-differences strategy in which we examine the *change* in the vote share. Let $\Delta(\text{Share Greens}_{me,2019-2015})$ denote the change in the Greens' vote share from 2015 to 2019 in municipality m for election e, which is given by

$$\Delta(\text{Share Greens}_{me,2019-2015}) = \theta_s + \tau_e + \beta \cdot \text{Participation Index}_{me} + \mu \mathbf{X}_{2m} + \xi_{me}. \quad (3.4)$$

Although first-differencing removes time-invariant municipality specific effects, we still include in some specifications a set of controls $\mathbf{X}_{2m}$ (potentially different to $\mathbf{X}_{1m}$ in Equation 3.3) to allow for time-varying effects of covariates.

In both Equations 3.3 and 3.4, all variables, dependent and explanatory alike, are standardized. β is the parameter of interest and captures the relationship between local strike participation rates and the (change in the) greens' vote share. The parameter reports the change in the outcome variable in standard deviation units due to a one standard deviation increase of the participation index. Observations are weighted by population figures from the Federal Statistical Office. For analyses at the municipality level, we calculate sandwiched standard error estimates allowing errors to be correlated across municipalities corresponding to the same county—i.e., clustered at the county level. For regressions at the county level, we use robust standard errors.

3.5.2 Main Results

Before presenting regression results, we show descriptive evidence on the variables of interest. Figure 3.6 contains maps displaying geographic patterns by grouping municipalities into quintiles according to the variables' distribution. Panel A shows the regional variation in the Green Party's vote share in the 2019 European elections. The Greens achieve their best results in urban areas of Western Germany, and with the exception of Berlin and three urban areas, the Greens are significantly less successful in Eastern Germany. The illustrated East-West divide is in line with official statistics. In West-Germany, the Greens obtained on average 22.31 percent of the votes, whereas an average 10.35 percent of voters elected the Greens in the East. Panel B shows the geographical variation of the *change* in the Greens' vote share between the 2015 and 2019 EU elections. Regions with the largest gains in votes can be found in the West and North of Germany as well as in large metropolitan regions. Panel C presents the spatial distribution of the standardized participation index at the time of the 2019 EU election. There is extensive regional variation with generally higher values in urban centers. The bivariate map in Panel D shows the correlation between changes in the Green Party's vote share and the protest participation index. To obtain the bivariate distribution, we divide municipalities into terciles of the two variables, resulting in a 3×3 matrix of possible outcomes (being in the lower, middle, or upper tercile, per variable). Intuitively, we assemble the bivariate map by overlaying the two univariate maps with the respective color scales, representing the terciles of the variable of interest. A more intense tinge of red indicates larger increases in the Green Party's vote share, whereas a more intense tinge of blue indicates higher protest participation rates. Geographies depicted in colors from the diagonal of the color matrix support the notion that electoral results are positively correlated with participation rates. For instance, low participation rates and small or even negative changes in the vote share of the Green Party are prevalent in Eastern Germany (light purple). In contrast, high levels of strike participation and large increases in vote shares for the Green Party are common in

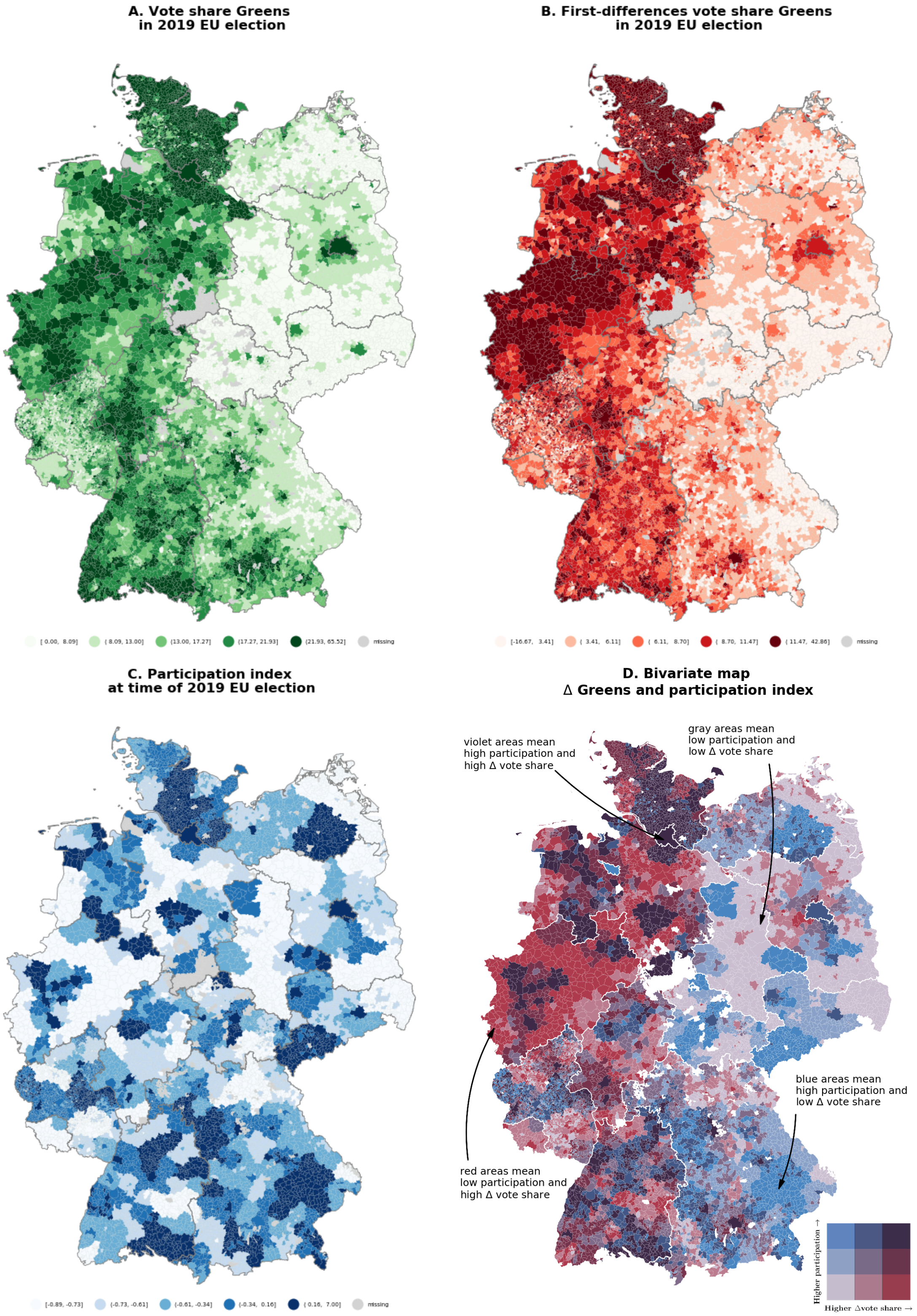


Figure 3.6. Spatial correlation of the vote share of the Greens and strike participation
Notes: The maps show, at the municipality level, (A) the vote share for the Greens in the 2019 election, (B) the first-difference of the vote share for the greens (2019-2015), (C) the strike participation index at the time of the EU election (standardized), and (D) the bivariate distribution of the first-differenced vote share for the greens and the participation index. The color scales in panels A-C correspond to quintiles. To generate the bivariate color scale in panel D, we blend the two univariate scales (in terciles, Δ Greens in red and participation index in blue) into one. Bold lines indicate state boundaries, thin lines represent municipality borders.

Table 3.2. Strike participation and the Greens' vote share

	(1) Greens 2019	(2) Δ Greens 2019-2015	(3) Δ Greens 2019-2015
Participation Index [std.]	0.1306***	0.0537**	0.0799**
	(0.0341)	(0.0270)	(0.0300)
Observations	12,189	12,189	455
Adjusted R^2	0.569	0.754	0.826
State FE	✓	✓	✓
Election FE	✓	✓	✓

Notes: The specifications use election results from the 2019 elections for the EU and the federal states of Brandenburg, Saxony, and Thuringia. The dependent variable is defined as the Greens' vote share, i.e. the number of votes relative to total votes cast. The explanatory variable is the participation index at the time of the elections. As all variables (dependent and explanatory) are standardized, population-weighted coefficients show the change in the outcome variable (in standard deviation units) due to a one standard deviation increase of the participation index. Columns 1 and 2 show results on the municipality level, while column 3 presents county-level results. All regressions include state and election fixed effects. Standard errors are clustered at the county level (number of clusters = 401) in columns 1 and 2. Robust standard errors are reported in column 3.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

North-Western Germany and Baden-Württemberg (dark purple). Areas with just a red or blue tinge are consequently not in line with our hypothesis.

Table 3.2 presents results from estimating Equations 3.3 and 3.4. In column 1, we estimate the relationship between cumulative FFF participation levels and the Green Party's vote share in 2019 at the municipality level. State-level fixed effects control for time-invariant factors that may have affected FFF participation or the Greens' vote share, such as political, economic, or cultural differences leading to different susceptibility levels for environmental topics between states. Election fixed effects account for the fact that electoral outcomes may vary across ballot types. For instance, individuals may choose to vote for different parties depending on the type of election. Our results indicate that participation in FFF climate strikes is predictive for the demand for green politics. To be precise, we find that a one-standard-deviation increase in our cumulative strike participation index is associated with an increase in the Green Party's vote share by .13 standard deviations. To eliminate unobserved time-invariant municipality specific effects, we estimate the relationship between our participation index and the *change* in the Greens' vote share between 2015 and 2019. Column 2 contains the estimate of this estimation and represents our preferred specification. A one-standard-deviation increase in the participation index is associated with an increase of .054 standard deviations. This increase in the standardized beta coefficient corresponds to an average gain of a quarter percentage point or 2.55 percent of the overall increase in the Green Party's vote share between 2015 and 2019. Finally, in column 3, we estimate the same specification as in column 2, but at the county level. Unlike the municipality-level measure of protest participation, the county-level measure has the advantage to not depend on weights (see section 3.2). Reassuringly, the coefficient is very similar to our preferred specification in column 2.

Table 3.3. Inclusion of controls

	(1) Baseline	(2) Income	(3) Unemployment	(4) Demographics
Participation Index [std.]	0.0537**	0.0439**	0.0560**	0.0422*
	(0.0270)	(0.0217)	(0.0269)	(0.0256)
Observations	12,189	12,185	12,066	12,189
Adjusted R^2	0.754	0.798	0.755	0.778
State FE	✓	✓	✓	✓
Election FE	✓	✓	✓	✓

Notes: The dependent variable is defined as the standardized change in Greens' vote share from 2015 to 2019. The baseline specification in column 1 corresponds to the specification in column 2 of Table 3.2. Each column adds a different set of control variables. Column 2 adds the logarithm of per capita income, column 3 adds the unemployment rate, and column 4 controls for population density (dummy: median split) and the age structure of the population. Clustered errors are reported in parentheses. See Table 3.2 for additional details.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.5.3 Additional Control Variables

Figure 3.6 displays considerable spatial variation in the Greens' vote share and participation in climate strikes, especially between East and West Germany. Therefore, one might be concerned that underlying factors, such as economic differences or differences in the age structure, which may explain the variation in both, the vote share of the Green Party and the participation in climate strikes, may confound the associations described above. For instance, municipalities with a high share of young people have a larger propensity to be more actively involved in school strikes for the climate and to vote 'green'. To check for the presence of potential confounders, we show results when including additional control variables to our preferred specification in Table 3.3. We include the logarithm of per capita income (column 2), the unemployment rate (column 3), and the age structure as well as population density (column 4).[16] It is reassuring to see that the coefficient size remains very similar throughout all specifications. The significance levels decrease slightly but remain below or just above the five percent level.

3.5.4 Voter Turnout as a Potential Mechanism

So far, the results have shown that the Green Party's vote share increased more in municipalities with higher levels of cumulative protest participation. The question arises whether this increase results from more people going to the ballot or a shift of votes from other parties to the Green Party. Concerning the former explanation, it is likely that the labeling of the European elections as 'climate election' may have mobilized additional voters who had abstained in previous elections. To probe this hypothesis, we investigate the relationship between strike participation and voter turnout in Table 3.4. Although voter turnout is not significantly correlated with climate strike participation, we find that the *change* in voter turnout is positively and significantly related to climate strike participation levels. We find that a one standard deviation increase in the participation index is associated with an

[16]Population density is an indicator variable that equals one for all municipalities with above-median population density.

Table 3.4. Strike participation and voter turnout

	(1) Turnout 2019	(2) Δ Turnout 2019-2015	(3) Δ Turnout 2019-2015
Participation Index [std.]	0.0140	0.0787***	0.1055***
	(0.0325)	(0.0181)	(0.0219)
Observations	12,189	12,189	455
Adjusted R^2	0.270	0.499	0.785
State FE	✓	✓	✓
Election FE	✓	✓	✓

Notes: The specifications use election results from the 2019 elections for the EU and the federal states of Brandenburg, Saxony, and Thuringia. The dependent variable is defined as the voter turnout, i.e. the number of votes relative to total votes cast. See Table 3.2 for additional details. Standard errors are clustered at the county level (number of clusters = 401) in columns 1 and 2. Robust standard errors are reported in column 3.
Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

increase of voter turnout by .08 standard deviations. This corresponds to half a percentage point increase in voter turnout, or equivalently to 3.5 percent of the increase in voter turnout from 2015 to 2019. We obtain similar results when estimating the specification at the county level.

Robustness Tests

We perform several sensitivity tests to assess the robustness of the findings. Overall, the sensitivity tests demonstrate that the main results are robust to other specifications of measuring climate strike participation, indicating that protest participation levels indeed correlate with electoral outcomes.

Cutoff Distances for the Climate Strike Participation Index.—In the process of constructing the climate strike participation index, we tried to be as flexible as possible. In principle, we allow every region to be a potential place of origin for any climate strike in Germany. In practice, however, participants are likely to come from nearby municipalities. This is also what we see in our data (see for instance Figure 3.4). In the first robustness test shown in Figure 3.7, we try to accommodate the insight about the locality of catchment areas for the individual strikes. To be precise, we limit the set of origin geographies based on the proximity to the strikes' location. The figure reports estimates and confidence intervals when estimating Equation 3.4 with the additional restriction of using only origins that are within a certain radius to the destination. The results show that choosing any cutoff of more than 20 kilometers will lead to positive and significant estimates that are similar to our baseline results presented earlier. Coefficient sizes decrease when using cutoffs that can realistically be seen as too small, i.e. of less than 17 kilometers. There is no significant correlation between strike participation and changes in the Greens' vote share when using these smaller distances.

Alternative Participation Measures.—The climate strike participation measure used throughout this paper is the residualized number of movements between origin-destination

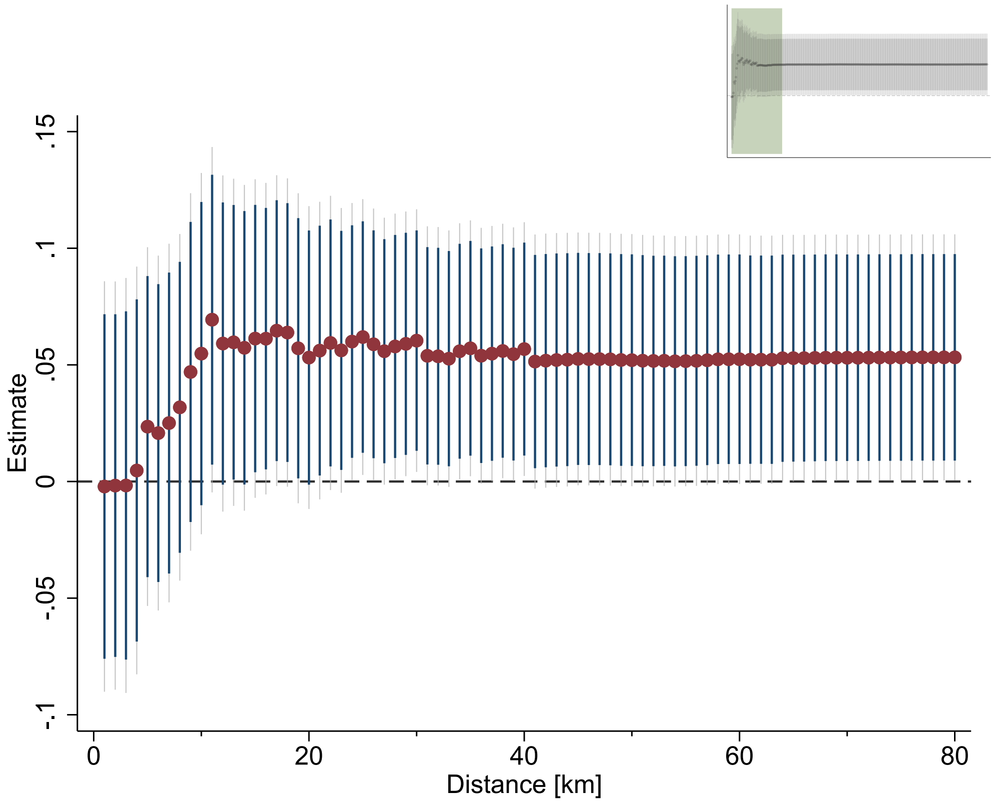


Figure 3.7. Cutoff distances for origins
Notes: The figure reports estimates and confidence intervals of Equation 3.4 (vertical axis) for different cutoff distances ranging from 0 to 80 kilometers (horizontal axis). The cutoff distance refers to the maximum distance until when counties are considered to be potential origins for a given strike. Gray bars represent 95% confidence intervals, while blue bars represent 90% confidence intervals, both arising from clustering standard errors at the county level.

pairs. We residualize journeys by removing the predictable variation using fixed effects for origin-destination pairs, the day of the week, the week of the year, and the month, as well as additional controls, such as weather and holidays. In Table 3.5, we show results when using even more stringent definitions by interacting origin-destination fixed effects with week- and month-fixed effects (column 2) and with week-, month-, and day-of-the-week fixed effects (column 3). Panel A shows results when using OLS, and Panel B presents estimates when using a Poisson model to account for the discrete nature of the dependent variable. The coefficient sizes remain similar to the baseline specification in column 1, while the standard errors increase slightly.

3.6 Concluding Remarks

This paper studies whether school strikes for climate, also known as Fridays for Future, have an impact on electoral outcomes in four elections in Germany in 2019, the peak of the FFF movement. To do so, we develop a novel strategy to identify the origin of climate protest participants exploiting cell phone-based tracking data. In a first step, we extract the excess number of journeys between origins and destinations, conditional on a large set of fixed effects and other controls. The residualized movements are matched with a

Table 3.5. Alternative participation measures

	(1)	(2)	(3)
	Interaction of ϑ_{ij} and γ_t		
	Baseline	Partial interaction	Fully interacted
Panel A: OLS			
Participation Index [std.]	0.0537**	0.0520*	0.0483*
	(0.0270)	(0.0272)	(0.0292)
Observations	12,189	12,189	12,189
Adjusted R^2	0.754	0.754	0.754
State FE	✓	✓	✓
Election FE	✓	✓	✓
Panel B: Poisson			
Participation Index [std.]	0.0541**	0.0469*	0.0501*
	(0.0251)	(0.0257)	(0.0300)
Observations	12,189	12,189	12,189
Adjusted R^2	0.754	0.754	0.754
State FE	✓	✓	✓
Election FE	✓	✓	✓

Notes: The dependent variable is defined as the standardized change in Greens' vote share from 2015 to 2019. The column header indicates how the process of residualizing journeys is varied. Column 1 uses the baseline specification presented in Equation 3.1. Column 2 further includes interactions of $\vartheta_{ij}\times$week and $\vartheta_{ij}\times$month. Column3 presents a fully interacted version (plus also including $\vartheta_{ij}\times$day-of-the-week). Panel A (B) contains estimates when estimating Equation 3.1 with OLS (Poisson). See Table 3.2 for additional details.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

self-constructed strike database, containing information on the exact time and location of more than 2,300 climate strikes. Consequently, we know for each strike the number of participants and the geographic distribution of their origins. In the second step, we find that this measure of local climate strike participation is related to electoral outcomes. To be precise, our results suggest that municipalities with higher strike participation rates have a larger propensity to vote for the Green Party. Furthermore, voter turnout is rising with increasing levels of climate strike participation. Our results survive a couple of sensitivity tests and are robust to including economic and demographic characteristics.

The results presented in this paper suggest that participation in climate protests may influence environmental awareness and political priorities. Previous literature in economics examines the transmission of values and political beliefs from parents to children. Our study presents first indications for a reverse transmission mechanism. The setting studied in this paper is interesting in this regard because the FFF movement is predominantly attended by school children, whereas parents are eligible to vote. The youth's participation in environmental protests may spark parents' concerns about climate change. While we do not have direct information on child-to-parents value transmissions, our results are indicative that children transmit their increased environmental awareness to their parents.

C Appendix to Chapter 3

C.1 Figures

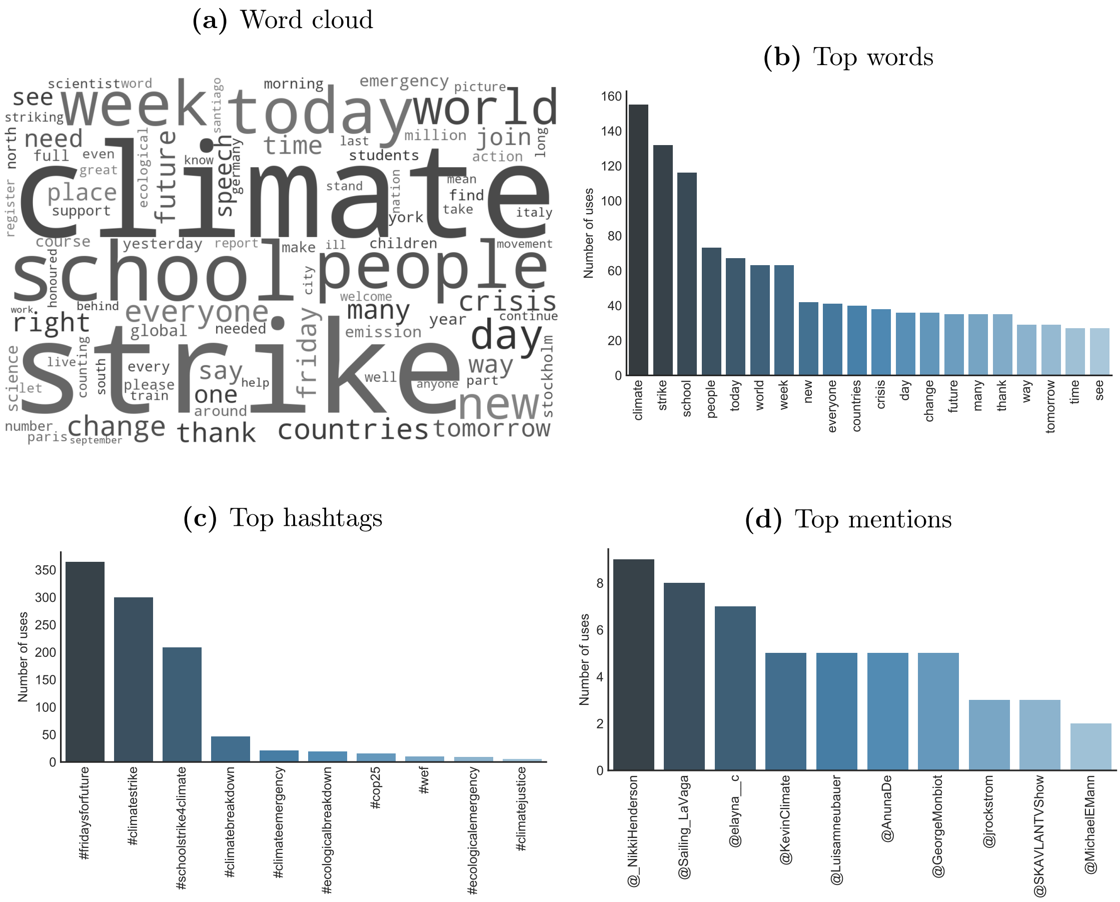


Figure C.1. Most widely used words, hashtags, and mentions in Greta Thunberg's tweets
Notes: Panel a displays a word cloud of Greta Thunberg's tweets in 2019. The larger a word is depicted, the more frequent it is used by Greta Thunberg in her tweets. The lowercased keywords are from the subset of words after removing urls, hashtags, mentions, digits, punctuation, and stopwords. Panels b, c, and d show the frequency of the most common words, hashtags, and mentions. The people behind the Twitter accounts are: @_NikkiHenderson, @Sailing_LaVaga, @elayna_c (sailors), @KevinClimate (Professor of energy and climate change), @Luisamneubauer (German climate activist), @AnunaDe (Belgian climate activist), @GeorgeMonbiot (British writer and political activist), @jrockstrom (Director of Potsdam Institute), @SKAVLANTVShow (Scandinavian talk show), @MichaelEMann (Professor of Atmospheric Science).
Source: Own representation with data from Twitter.

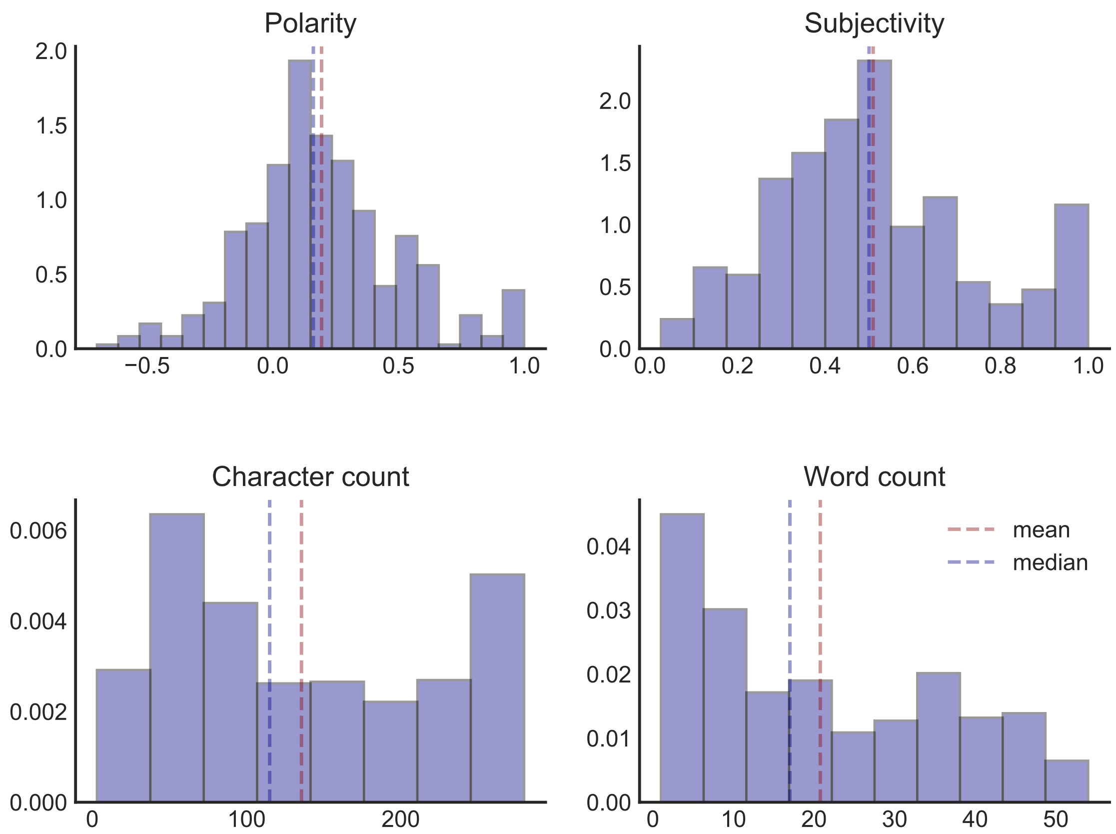


Figure C.2. Sentiment and length of Greta Thunberg's tweets
Notes: The two top panels show the conditional distribution of the tweets' sentiment. Greta Thunberg's tweets tend to be more positively formulated (polarity values larger than zero), with the right balance between objective and subjective language. Her tweets contain on average 135 characters and a bit more than 20 words.
Source: Own representation with data from Twitter.

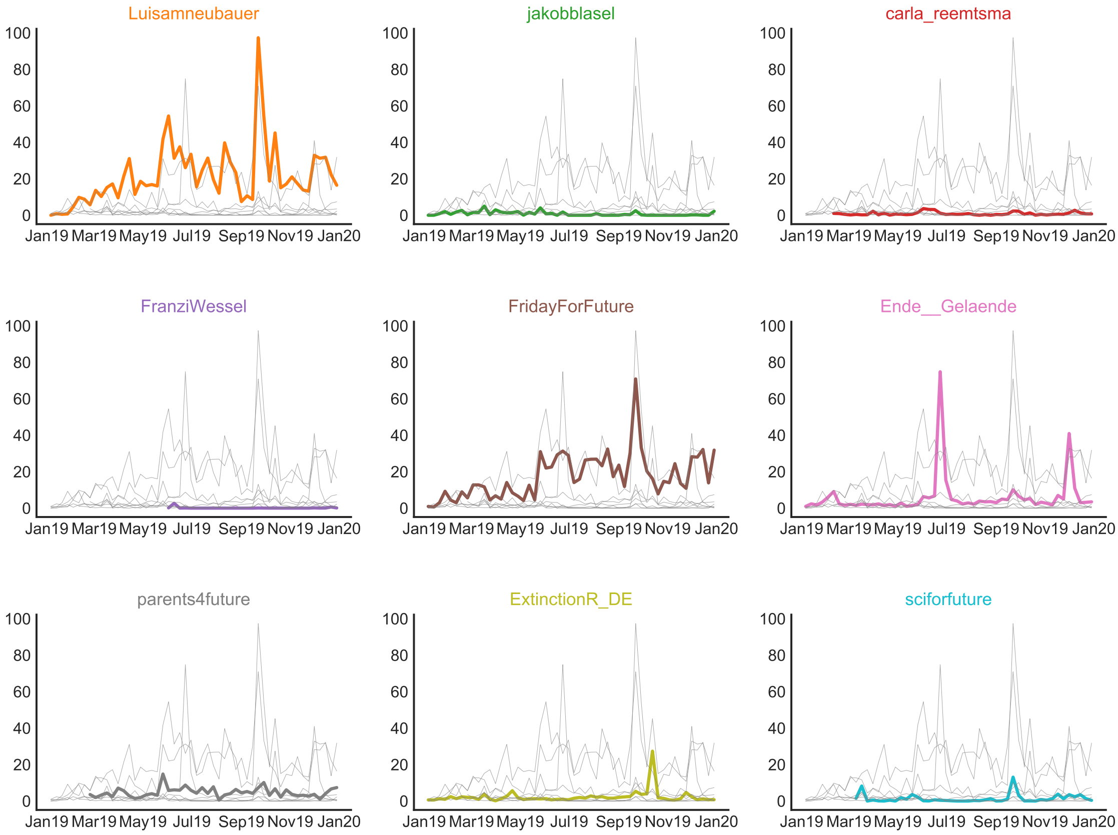


Figure C.3. The twitter feed of influential German FFF activists

Notes: The figure plots the weekly number of likes [in thousand] of influential German climate activists and groups over the year 2019.

Source: Own representation with data from Twitter.

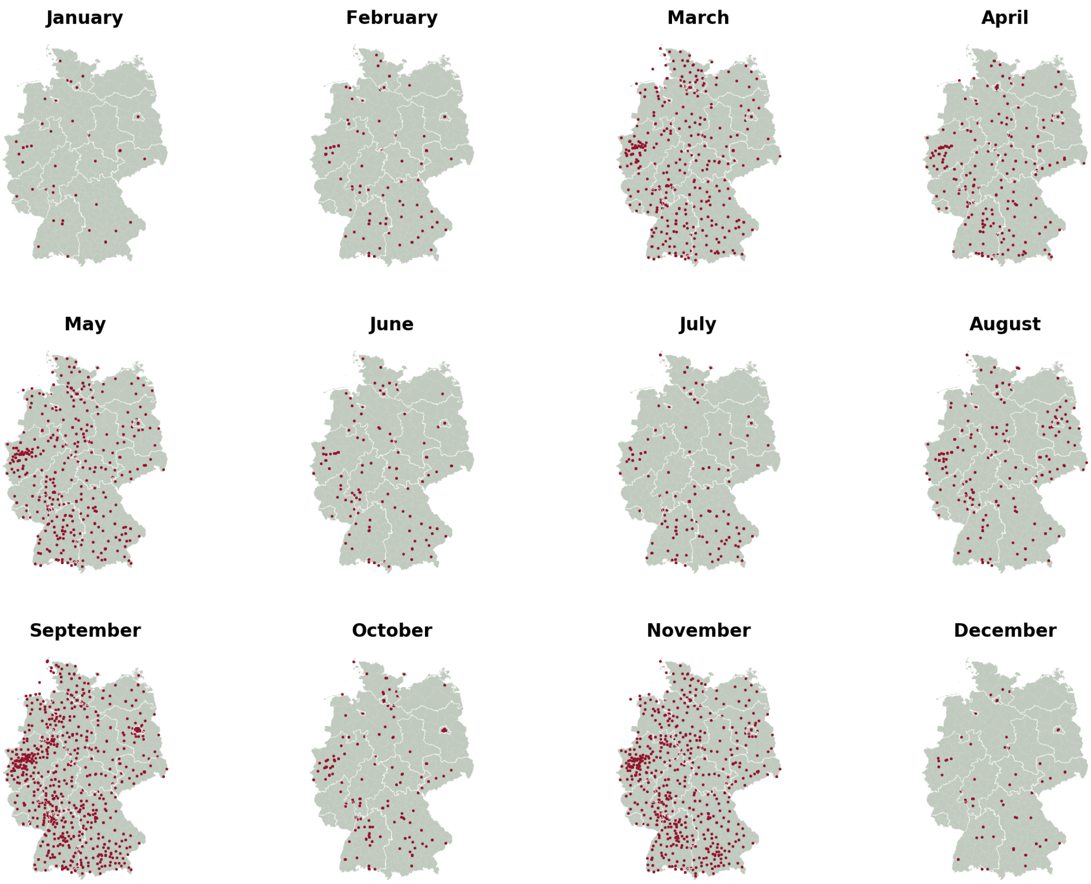


Figure C.4. Strikes in 2019, variation across months

Notes: The maps show the climate strikes (red dots) in our data base over the year 2019. The bold white lines indicate state boundaries and the thin white lines represent county boundaries.

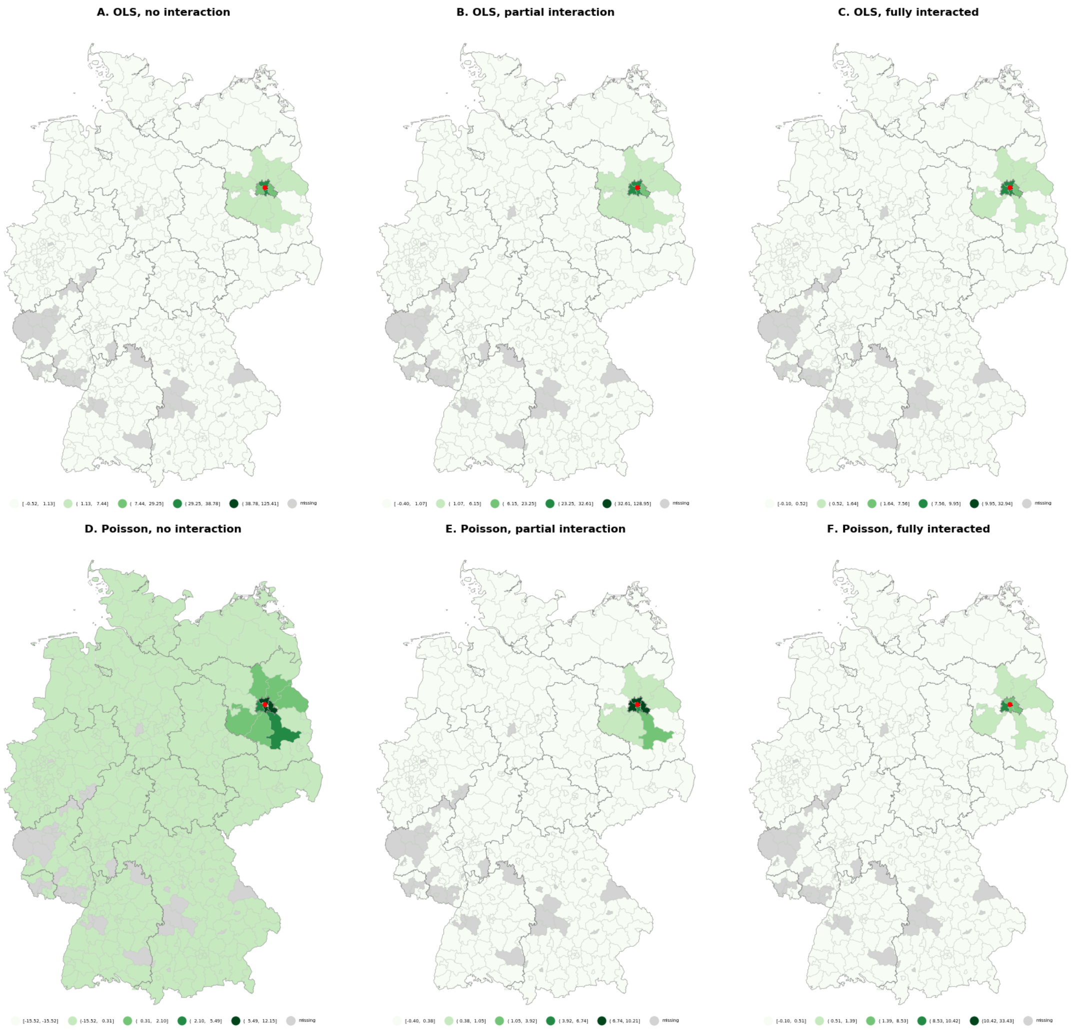


Figure C.5. Alternative participation measures for the climate strike in Berlin

Notes: The maps show residualized movements (in thousand) for the climate strike in Berlin (March, 29). Column 1 uses the baseline specification in Equation 3.1, column 2 further adds $\vartheta_{ij} \times \eta_w$ and $\vartheta_{ij} \times \psi_m$, and column 3 presents a fully interacted version (plus also including $\vartheta_{ij} \times \varphi_d$). The top row uses OLS to estimate Equation 3.1, while the bottom row shows results when using Poisson. A darker shade of green indicates that more people were coming to the climate strike conditional on the controls discussed in the text. The color scale classification is obtained by using the Fisher-Jenks natural breaks algorithm. See Figure 3.4 for additional details.

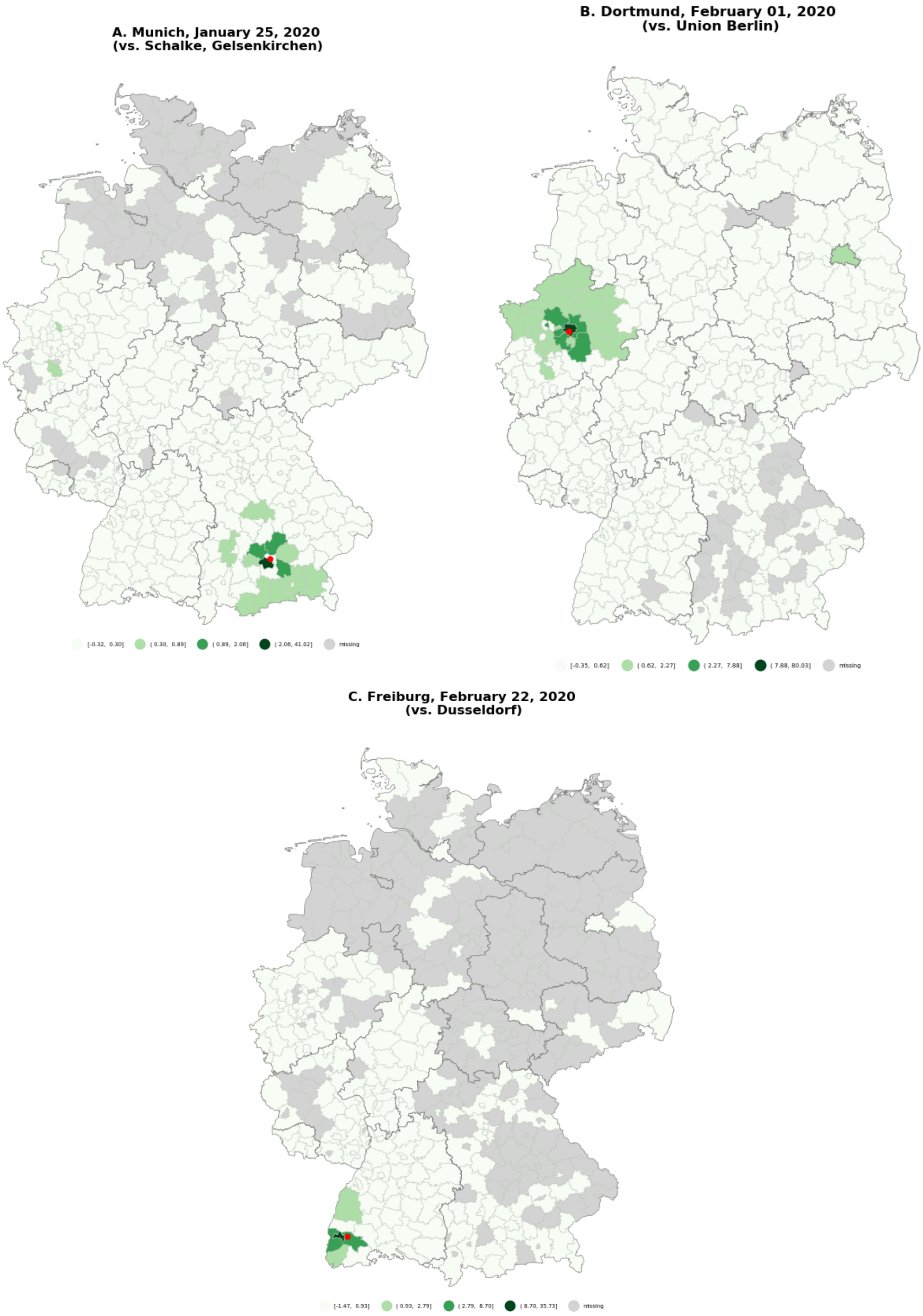


Figure C.6. Validation: Attendance at soccer games

Notes: The maps show residualized movements (in thousand) for soccer matches in Munich, Dortmund, and Freiburg. A darker shade of green indicates that more people were coming to the soccer matches conditional on the controls discussed in the text. The color scale classification is obtained by using the Fisher-Jenks natural breaks algorithm. The red dots mark the games' location, gray areas indicate missing data (censored), bold gray lines show state boundaries, and thin gray lines represent the regions defined by *Teralytics*.

Bibliography

90MIN.DE (2020). List of largest rivalries in german soccer. https://www.90min.de/posts/3729551-purer-hass-die-15-groessten-fan-feindschaften-deutschlands, [Online; accessed 29-July-2020].

ACEMOGLU, D., HASSAN, T. A. and TAHOUN, A. (2018). The Power of the Street: Evidence from Egypt's Arab Spring. *The Review of Financial Studies* (1), 1–42.

AHAMMER, A., HALLA, M. and SCHNEEWEIS, N. (2020). The effect of prenatal maternity leave on short and long-term child outcomes. *Journal of Health Economics*, 102250.

AIZER, A., ELI, S., FERRIE, J. and LLERAS-MUNEY, A. (2016). The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review* (4), 935–71.

AKEE, R., COPELAND, W., COSTELLO, E. J. and SIMEONOVA, E. (2018). How does household income affect child personality traits and behaviors? *American Economic Review* (3), 775–827.

ALBAGLI, P. and RAU, T. (2018). The Effects of a Maternity Leave Reform on Children's Abilities and Maternal Outcomes in Chile. *The Economic Journal*.

ALMOND, D. and CURRIE, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives* (3), 153–72.

—, — and DUQUE, V. (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature* (4), 1360–1446.

ANGRIST, J., AZOULAY, P., ELLISON, G., HILL, R. and LU, S. F. (2017). Economic Research Evolves: Fields and Styles. *American Economic Review* (5), 293–97.

— and PISCHKE, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Tech. rep., Princeton University Press.

— and — (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives* (2), 3–30.

ATHEY, S. (2018). The Impact of Machine Learning on Economics. In *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, pp. 507–47.

— and IMBENS, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives* (2), 3–32.

AVDIC, D. and KARIMI, A. (2018). Modern Family? Paternity Leave and Marital Stability. *American Economic Journal: Applied Economics* (4), 283–307.

Baily, C. (2009). Reverse Intergenerational Learning: A Missed Opportunity? *AI & Society* (1), 111–115.

Baker, M. and Milligan, K. (2008). Maternal employment, breastfeeding, and health: Evidence from maternity leave mandates. *Journal of Health Economics* (4), 871–887.

Bandura, A. (1973). *Aggression: A social learning analysis.* Prentice-Hall.

— (2007). *Social learning theory of aggression.* In J. F. Knutston (Ed.), The control of aggression: Implications from basic research (pp.201–252). New Brunswick, NJ: Transaction Publishers.

Barker, D. J. (1990). The fetal and infant origins of adult disease. *BMJ* (6761), 1111–1111.

Baum, C. L. (2003). The effect of state maternity leave legislation and the 1993 Family and Medical Leave Act on employment and wages. *Labour Economics* (5), 573–596.

Bender, S. and Hirschenauer, F. (1993). Regionale Unterschiede in der Frauenerwerbstätigkeit – Eine Typisierung westdeutscher Arbeitsmarktregionen. *Mitteilungen aus der Arbeitsmarkt-und Berufsforschung* (3), 294–312.

Berger, L. M., Hill, J. and Waldfogel, J. (2005). Maternity leave, early maternal employment and child health and development in the US. *The Economic Journal* (501), F29–F47.

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin* (1), 59.

Beuchert, L. V., Humlum, M. K. and Vejlin, R. (2016). The length of maternity leave and family health. *Labour Economics*, 55–71.

Bisin, A. and Verdier, T. (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory* (2), 298–319.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics* (2), 455–467.

Blau, F. D. and Kahn, L. M. (2013). Female Labor Supply: Why Is the United States Falling Behind? *American Economic Review* (3), 251–56.

Branscombe, N. R. and Wann, D. L. (1992). Role of Identification with a Group, Arousal, Categorization Processes, and Self-Esteem in Sports Spectator Aggression. *Human Relations* (10), 1013–1033.

Buckles, K. S. and Hungerman, D. M. (2013). Season of birth and later outcomes: Old questions, new answers. *The Review of Economics and Statistics* (3), 711–724.

Bullinger, L. R. (2019). The Effect of Paid Family Leave on Infant and Parental Health in the United States. *Journal of Health Economics*, 101–116.

Bursztyn, L., Cantoni, D., Yang, D. Y., Yuchtman, N. and Zhang, Y. J. (forthcoming). Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements. *American Economic Review: Insights.*

Bütikofer, A., Riise, J. and Skira, M. (forthcoming). The Impact of Paid Maternity Leave on Maternal Health. *American Economic Journal: Economic Policy*.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Canetti, L., Bachar, E., Galili-Weisstub, E., De-Nour, A. and Shalev, A. (1997). Parental bonding and mental health in adolescence. *Adolescence* (126), 381.

Cantoni, D., Hagemeister, F. and Westcott, M. (2020). Persistence and Activation of Right-Wing Political Ideology. *mimeo*.

—, Yang, D. Y., Yuchtman, N. and Zhang, Y. J. (2019). Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement. *The Quarterly Journal of Economics* (2), 1021–1077.

Card, D. and Dahl, G. B. (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *The Quarterly Journal of Economics* (1), 103–143.

Carneiro, P., Løken, K. V. and Salvanes, K. G. (2015). A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children. *Journal of Political Economy* (2), 365–412.

Center on the Developing Child at Harvard University (2016). From Best Practices to Breakthrough Impacts: A Science-Based Approach to Building a More Promising Future for Young Children and Families. https://developingchild.harvard.edu/.

Central Sports Intelligence Unit (2018). Annual report for 2017/18 season. *Zentrale Informationsstelle Sporteinsätze (ZIS), Landesamt für Zentrale Polizeiliche Dienste Nordrhein-Westfalen*.

Chatterji, P. and Markowitz, S. (2005). Does the Length of Maternity Leave Affect Maternal Health? *Southern Economic Journal*, 16–41.

—, — and Brooks-Gunn, J. (2013). Effects of early maternal employment on maternal health and well-being. *Journal of Population Economics* (1), 285–301.

Coase, R. H. (1960). The problem of social cost. In *Classic papers in natural resource economics.*, Palgrave Macmillan, London, pp. 87–137.

Cohen, M., Rust, R., Steen, S. and Tidd, S. (2004). Willingness-to-pay for crime control programs. *Criminology* (1), 89–110.

Cook, P. J. and Durrance, C. P. (2013). The virtuous tax: Lifesaving and crime-prevention effects of the 1991 federal alcohol-tax increase. *Journal of Health Economics* (1), 261–267.

Currie, J. and Almond, D. (2011). Human capital development before age five. *Handbook of Labor Economics*, 1315–1486.

— and Schwandt, H. (2013). Within-mother analysis of seasonal patterns in health at birth. *Proceedings of the National Academy of Sciences* (30), 12265–12270.

Dahl, G. and DellaVigna, S. (2009). Does movie violence increase violent crime? *The Quarterly Journal of Economics* (2), 677–734.

Dahl, G. B. and Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review* (5), 1927–1956.

—, Løken, K. V., Mogstad, M. and Salvanes, K. V. (2016). What is the case for paid maternity leave? *Review of Economics and Statistics*.

Damerell, P., Howe, C. and Milner-Gulland, E. J. (2013). Child-orientated Environmental Education Influences Adult Knowledge and Household Behaviour. *Environmental Research Letters* (1), 015016.

Danzer, N., Halla, M., Schneeweis, N. E. and Zweimüller, M. (forthcoming). Parental Leave, (In)formal Childcare and Long-term Child Outcomes. *Journal of Human Resources*.

— and Lavy, V. (2017). Paid Parental Leave and Children's Schooling Outcomes. *The Economic Journal* (608), 81–117.

Daumann, F. (2012). Sollen Vereine die Kosten für Polizeieinsätze im professionellen Fußball tragen? *Wirtschaftliche Freiheit: Das ordnungspolitische Journal*.

Dave, D., McNichols, D. and Sabia, J. J. (2020). The Contagion Externality of a Superspreading Event: The Sturgis Motorcycle Rally and COVID-19. *Southern Economic Journal*.

De Moor, J., Uba, K., Wahlström, M., Wennerhag, M. and De Vydt, M. (2020). Protest for a future II: Composition, mobilization and motives of the participants in Fridays For Future climate protests on 20-27 September, 2019, in 19 cities around the world.

Deloitte (2019). World in motion - Annual Review of Football Finance 2019. *Sports Business Group*.

Depetris-Chauvin, E., Durante, R. and Campante, F. (2020). Building Nations through Shared Experiences: Evidence from African Football. *American Economic Review* (5), 1572–1602.

derbys.org (2020). List of socer derbies. https://www.derbys.org/fussballderbies/deutschland/, [Online; accessed 29-July-2020].

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H. and Sears, R. R. (1939). *Frustration and aggression*. Yale University Press.

Dubourg, R., Hamed, J., Thorns, J. *et al.* (2005). The economic and social costs of crime against individuals and households 2003/04. *Home Office online report* (05).

Dustmann, C. and Schönberg, U. (2012). Expansions in Maternity Leave Coverage and Children' s Long-Term Outcomes. *American Economic Journal: Applied Economics* (3), 190–224.

Einav, L. and Levin, J. (2014). Economics in the age of big data. *Science* (6210).

Ekberg, J., Eriksson, R. and Friebel, G. (2013). Parental leave - A policy evaluation of the Swedish "Daddy-Month" reform. *Journal of Public Economics* (1), 131–143.

Enns, M. W., Cox, B. J. and Clara, I. (2002). Parental bonding and adult psychopathology: results from the US National Comorbidity Survey. *Psychological Medicine* (6), 997–1008.

Fabel, M. (2021). Maternity leave and children's health outcomes in the long-term. *Journal of Health Economics* (76), 102431.

Falk, A. and Kosse, F. (2016). Early Childhood Environment, Breastfeeding and the Formation of Preferences. *HCEO Working Paper* (036).

Federal Administrative Court (2019). Bremen police fee for high-risk events is in principle lawful. *Press release* (26/2019).

Federal Ministry of Finance (2020). Data on the german federal budget. https://www.bundeshaushalt.de/#, [Online; accessed 29-March-2020].

Federal Statistical Office (1981). *Statistisches Jahrbuch für die Bundesrepublik Deutschland 1981*. Statistisches Bundesamt (Wiesbaden).

Federal Statistical Office (2012). Diagnosedaten der Patienten und Patientinnen in Krankenhäusern (einschl. Sterbe-und Stundenfälle). *Fachserie 12 Gesundheit* (6.2.1).

Federal Statistical Office (2015). Data on the cost of illness by disease and age group. https://www.destatis.de/EN/Themes/Society-Environment/Health/Cost-Illness/Tables/disease-categories-age.html, [Online; accessed 22-April-2020].

Feltes, T. (2010). Fußballgewalt als misslungene Kommunikation - Lösungsansätze abseits von Repression. *Neue Praxis* (4).

Fernández, R. and Fogli, A. (2009). Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *American Economic Journal: Macroeconomics* (1), 146–77.

—, Fogli, A. and Olivetti, C. (2004). Mothers and Sons: Preference Formation and Female Labor Force Dynamics. *The Quarterly Journal of Economics* (4), 1249–1299.

Figlio, D., Giuliano, P., Özek, U. and Sapienza, P. (2019). Long-Term Orientation and Educational Performance. *American Economic Journal: Economic Policy* (4), 272–309.

Finkel, S. E. and Muller, E. N. (1998). Rational Choice and the Dynamics of Collective Political Action: Evaluating Alternative Models with Panel Data. *The American Political Science Review* (1), 37–49.

— and Opp, K.-D. (1991). Party Identification and Participation in Collective Political Action. *The Journal of Politics* (2), 339–371.

Flurry, L. A. and Burns, A. C. (2005). Children's Influence in Purchase Decisions: a Social Power Theory Approach. *Journal of Business Research* (5), 593–601.

Forschungsgruppe Wahlen e.V. (2019). Politbarometer Juni II 2019. https://www.forschungsgruppe.de/Umfragen/Politbarometer/Archiv/Politbarometer-2019/Juni-II-2019/, [Online; accessed 12-February-2021].

Frech, A. and Kimbro, R. T. (2011). Maternal Mental Health, Neighborhood Characteristics, and Time Investments in Children. *Journal of Marriage and Family* (3), 605–620.

Friedman, N. J. and Zeiger, R. S. (2005). The role of breast-feeding in the development of allergies and asthma. *Journal of Allergy and Clinical Immunology* (6), 1238 – 1248.

Gans, J. S. and Leigh, A. (2009). Born on the first of July: An (un) natural experiment in birth timing. *Journal of Public Economics* (1), 246–263.

Gentzkow, M., Kelly, B. and Taddy, M. (2019). Text as Data. *Journal of Economic Literature* (3), 535–74.

Ginja, R., Jans, J. and Karimi, A. (2020). Parental Leave Benefits, Household Labor Supply and Children's Long-run Outcomes. *Journal of Labor Economics* (1), 261–320.

Glaubitz, C., Steglich, F., Koch, M., Klodt, H., Klatt, T., Hausmann, B. and Bliesener, T. (2016). The costs of youth crime in Germany - An empirical contribution. *Monatsschrift für Kriminologie und Strafrechtsreform* (2), 123–139.

Hagedorn, G., Kalmus, P., Mann, M., Vicca, S., Van den Berge, J., van Ypersele, J.-P., Bourg, D., Rotmans, J., Kaaronen, R., Rahmstorf, S., Kromp-Kolb, H., Kirchengast, G., Knutti, R., Seneviratne, S. I., Thalmann, P., Cretney, R., Green, A., Anderson, K., Hedberg, M., Nilsson, D., Kuttner, A. and Hayhoe, K. (2019). Concerns of young protesters are justified. *Science* (6436), 139–140.

Hamermesh, D. S. (2013). Six Decades of Top Economics Publishing: Who and How? *Journal of Economic Literature* (1), 162–72.

Hank, K. and Kreyenfeld, M. (2001). Childcare and fertility in (Western) Germany. *Max Planck Institute for Demographic Research Working Paper* (19).

Hener, T. (2019). Noise Pollution and Violent Crime. *mimeo*.

Hersch, J. and Viscusi, W. K. (2006). The Generational Divide in Support for Environmental Policies: European Evidence. *Climatic Change* (1), 121–136.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association* (396), 945–960.

Hoynes, H., Miller, D. and Simon, D. (2015). Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy* (1), 172–211.

Huber, M. (2019). An introduction to flexible methods for policy evaluation.

Huebener, M., Kuehnle, D. and Spiess, C. K. (2019). Parental leave policies and socio-economic gaps in child development: Evidence from a substantial benefit reform using administrative data. *Labour Economics*.

Imbens, G. W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature* (4), 1129–79.

Jenks, G. F. (1967). The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, 186–190.

Jürges, H. and Schneider, K. (2011). Why Young Boys Stumble: Early Tracking, Age and Gender Bias in the German School System. *German Economic Review* (4), 371–394.

Kluve, J. and Tamm, M. (2013). Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: evidence from a natural experiment. *Journal of Population Economics* (3), 983–1005.

Konsortium Bildungsberichterstattung (2006). *Bildung in Deutschland: ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bertelsmann, on behalf of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany and the Federal Ministry of Education and Research.

Kotsadam, A. and Finseraas, H. (2011). The state intervenes in the battle of the sexes: Causal effects of paternity leave. *Social Science Research* (6), 1611–1622.

Kovanen, L., Kaski, K., Kertész, J. and Saramäki, J. (2013). Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proceedings of the National Academy of Sciences* (45), 18070–18075.

Kreindler, G. E. and Miyauchi, Y. (2021). *Measuring Commuting and Economic Activity inside Cities with Cell Phone Records*. Working Paper 28516, National Bureau of Economic Research.

Lalive, R., Schlosser, A., Steinhauer, A. and Zweimüller, J. (2014). Parental leave and mothers' careers: The relative importance of job protection and cash benefits. *Review of Economic Studies* (1), 219–265.

— and Zweimüller, J. (2009). How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments. *The Quarterly Journal of Economics* (3), 1363–1402.

Lange, C., Schenk, L. and Bergmann, R. (2007). Distribution, duration and temporal trend of breastfeeding in Germany. Results of the German Health Interview and Examination Survey for Children and Adolescents (KiGGS). *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* (5-6), 624–633.

Langedijk, S., Vollbracht, I. and Paruolo, P. (2019). *The Potential of Administrative Microdata for Better Policy-Making in Europe*, Cham: Springer International Publishing, pp. 333–346.

LaSala, M. C. (2000). Lesbians, Gay Men, and their Parents: Family Therapy for the Coming-out Crisis. *Family Process* (1), 67–81.

Lawson, D. F., Stevenson, K. T., Peterson, M. N., Carrier, S. J., Strnad, R. L. and Seekamp, E. (2019). Children Can Foster Climate Change Concern Among their Parents. *Nature Climate Change* (6), 458–462.

Le Huërou-Luron, I., Blat, S. and Boudry, G. (2010). Breast-v. formula-feeding: impacts on the digestive tract and immediate and long-term health effects. *Nutrition Research Reviews* (1), 23–36.

Leamer, E. E. (1983). Let's Take the Con out of Econometrics. *The American Economic Review* (1), 31–43.

Lindo, J. M., Siminski, P. and Swensen, I. D. (2018). College party culture and sexual assault. *American Economic Journal: Applied Economics* (1), 236–65.

Lodge, C., Tan, D., Lau, M., Dai, X., Tham, R., Lowe, A., Bowatte, G., Allen, K. and Dharmage, S. (2015). Breastfeeding and asthma and allergies: a systematic review and meta-analysis. *Acta Paediatrica* (S467), 38–53.

Lupien, S. J., McEwen, B. S., Gunnar, M. R. and Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience* (6), 434.

Madestam, A., Shoag, D., Veuger, S. and Yanagizawa-Drott, D. (2013). Do Political Protests Matter? Evidence from the Tea Party Movement. *The Quarterly Journal of Economics* (4), 1633–1685.

Marie, O. (2016). Police and thieves in the stadium: measuring the (multiple) effects of football matches on crime. *Journal of the Royal Statistical Society. Series A: Statistics in Society* (1), 273–292.

Mause, K. (2020). Football matches, police operations and taxpayers: economic remarks on the police-cost debate. *List Forum* (45), 423–440.

McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences* (1), 33–44.

Miller, T., Cohen, M. and Wiersema, B. (1996). *Victim Costs and Consequences: A New Look*. US Department of Justice.

Milligan, K. and Stabile, M. (2011). Do Child Tax Benefits Affect the Well-Being of Children? Evidence from Canadian Child Benefit Expansions. *American Economic Journal: Economic Policy* (3), 175–205.

Montolio, D. and Planells-Struse, S. (2016). How time shapes crime: The temporal impacts of football matches on crime. *Regional Science and Urban Economics*, 99–113.

— and — (2019). Measuring the negative externalities of a private leisure activity: hooligans and pickpockets around the stadium. *Journal of Economic Geography* (2), 465–504.

Morrill, M. S. (2011). The effects of maternal employment on the health of school-age children. *Journal of Health Economics* (2), 240–257.

Mullainathan, S. and Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives* (2), 87–106.

Murphy, E., Wickramaratne, P. and Weissman, M. (2010). The stability of parental bonding reports: A 20-year follow-up. *Journal of Affective Disorders* (1), 307 – 315.

Nassauer, A. (2011). From Hate to Collective Violence: Research and Practical Implications. *Journal of Hate Studies* (1), 199–220.

Neugart, M. and Ohlsson, H. (2013). Economic incentives and the timing of births: evidence from the german parental benefit reform of 2007. *Journal of Population Economics* (1), 87–108.

Oddy, W. H., Kendall, G. E., Li, J., Jacoby, P., Robinson, M., de Klerk, N. H., Silburn, S. R., Zubrick, S. R., Landau, L. I. and Stanley, F. J. (2010). The Long-Term Effects of Breastfeeding on Child and Adolescent Mental Health: A Pregnancy Cohort Study Followed for 14 Years. *The Journal of Pediatrics* (4), 568–574.

OECD (2020). Data on length of maternity leave, parental leave and paid father-specific leave. https://stats.oecd.org/index.aspx?queryid=54760, [Online; accessed 30-December-2020].

Ondrich, J., Spiess, K. C. and Yang, Q. (2002). The Effect of Maternity Leave on Women's Pay in Germany 1984-1994. *DIW Discussion Paper* (289).

Onnela, J.-P., Arbesman, S., González, M. C., Barabási, A.-L. and Christakis, N. A. (2011). Geographic constraints on social network groups. *PLoS one* (4), e16939.

—, Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J. and Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences* (18), 7332–7336.

Papp, L. M. (2014). Longitudinal associations between breastfeeding and observed mother–child interaction qualities in early childhood. *Child: Care, Health and Development* (5), 740–746.

Patel, V., Rahman, A., Jacob, K. S. and Hughes, M. (2004). Effect of maternal mental health on infant growth in low income countries: new evidence from South Asia. *BMJ* (7443), 820–823.

Pilz, G. A. (2005). Vom Kuttenfan und Hooligan zum Ultra und Hooltra–Wandel des Zuschauerverhaltens im Profifußball. *Deutsche Polizei* (11), 6–12.

Police Crime Statistics (2018). Report 2018, abrigded version. *Federal Criminal Police Office.*

Politbarometer (2019). Cumulated Data Set, Forschungsgruppe Wahlen, Mannheim. *GESIS Data Archive, Cologne* (ZA7599), Data file Version 1.0.0.

Poutvaara, P. and Priks, M. (2009). The effect of police intelligence on group violence: Evidence from reassignments in Sweden. *Journal of Public Economics* (3-4), 403–411.

PwC (2016). Survey among soccer fans ('Fußballfan-Befragung').

Räikkönen, K., Pesonen, A.-K., Roseboom, T. J. and Eriksson, J. G. (2012). Early determinants of mental health. *Best Practice & Research Clinical Endocrinology & Metabolism* (5), 599–611.

Rees, D. I. and Schnepel, K. T. (2009). College Football Games and Crime. *Journal of Sports Economics* (1), 68–87.

Roper, T. and Thompson, A. (2006). Estimating the costs of crime in New Zealand in 2003/04. *New Zealand Treasury Working Paper.*

Ruhm, C. J. (2000). Parental leave and child health. *Journal of Health Economics* (6), 931–960.

Sayour, N. (2019). The impact of maternal care on child development: Evidence from sibling spillover effects of a parental leave expansion. *Labour Economics*, 167–186.

Schönberg, U. and Ludsteck, J. (2014). Expansions in Maternity Leave Coverage and Mothers' Labor Market Outcomes after Childbirth. *Journal of Labor Economics* (3), 469–505.

Selbmann, H. and Thieme, C. (1988). Die Bayerische Perinatalerhebung im Jahre 1987. *Bayerisches Ärzteblatt* (8), 297–304.

Shonkoff, J., Boyce, W. and McEwen, B. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities: Building a new framework for health promotion and disease prevention. *JAMA* (21), 2252–2259.

Smith, E. K. and Bognar, J. (2019). A window for climate action. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-65376-7, [Online; accessed 10-February-2021].

Sommer, M., Rucht, D., Haunss, S. and Zajak, S. (2019). Fridays for Future: Profil, Entstehung und Perspektiven der Protestbewegung in Deutschland. *ipb working paper* (2/2019).

Spiegel (2020). Data on friends and rivals of german soccer clubs. https://www.spiegel.de/sport/fussball/fussball-wer-sind-die-freunde-und-rivalen-ihres-lieblingsvereins-a-1219354.html, [Online; accessed 29-July-2020].

Statista (2020). Marktanteile der einzelnen Netzbetreiber an den Mobilfunkanschlüssen in Deutschland von 1998 bis 2020. https://de.statista.com/statistik/daten/studie/3028/umfrage/marktanteile-der-netzbetreiber-am-mobilfunkmarkt-in-deutschland-seit-1998/, [Online; accessed 13-February-2021].

Stearns, J. (2015). The effects of paid maternity leave: Evidence from Temporary Disability Insurance. *Journal of Health Economics*, 85–102.

The Economist (2019). The Greta effect. https://www.economist.com/graphic-detail/2019/08/19/the-greta-effect, [Online; accessed 11-February-2021].

The Guardian (2019a). Across the globe, millions join biggest climate protest ever. https://www.theguardian.com/environment/2019/sep/21/across-the-globe-millions-join-biggest-climate-protest-ever, [Online; accessed 11-February-2021].

The Guardian (2019b). 'Greta Thunberg effect' driving growth in carbon offsetting. https://www.theguardian.com/environment/2019/nov/08/greta-thunberg-effect-driving-growth-in-carbon-offsetting, [Online; accessed 11-February-2021].

Time (2019). Students From 1,600 Cities Just Walked Out of School to Protest Climate Change. It Could Be Greta Thunberg's Biggest Strike Yet. https://time.com/5595365/global-climate-strikes-greta-thunberg/, [Online; accessed 11-February-2021].

Tomalski, P., Moore, D. G., Ribeiro, H., Axelsson, E. L., Murphy, E., Karmiloff-Smith, A., Johnson, M. H. and Kushnerenko, E. (2013). Socioeconomic status and functional brain development - associations in early infancy. *Developmental Science* (5), 676–687.

Tronick, E. and Reck, C. (2009). Infants of Depressed Mothers. *Harvard Review of Psychiatry* (2), 147–156.

Van der Klaauw, B. (2014). From micro data to causality: Forty years of empirical labor economics. *Labour Economics*, 88–97.

Victora, C. G., Bahl, R., Barros, A. J., França, G. V., Horton, S., Krasevec, J., Murch, S., Sankar, M. J., Walker, N., Rollins, N. C. *et al.* (2016). Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *The Lancet* (10017), 475–490.

—, Horta, B. L., de Mola, C. L., Quevedo, L., Pinheiro, R. T., Gigante, D. P., Gonçalves, H. and Barros, F. C. (2015). Association between breastfeeding and intelligence, educational attainment, and income at 30 years of age: a prospective birth cohort study from Brazil. *The Lancet Global Health* (4), e199–e205.

Wann, D. L. and Branscombe, N. R. (1993). Sports fans: Measuring degree of identification with their team. *International Journal of Sport Psychology*.

Warren, M. (2019). Thousands of scientists are backing the kids striking for climate change. *Nature* (7748), 291–293.

Wicker, P., Whitehead, J. C., Johnson, B. K. and Mason, D. S. (2017). The Effect of Sporting Success and Management Failure On Attendance Demand In The Bundesliga: A Revealed and Stated Preference Travel Cost Approach. *Applied Economics* (52), 5287–5295.

Zeit Online (2019). Hunderttausende demonstrieren für das Klima. https://www.zeit.de/gesellschaft/zeitgeschehen/2019-11/fridays-for-future-klimaprotest-demonstrationen-klimaschutz, [Online; accessed 11-February-2021].

Zmarzlik, J., Zipperer, M. and Viethen, H. P. (1999). *Mutterschutzgesetz, Mutterschaftsleistungen, Bundeserziehungsgeldgesetz: mit Mutterschutzverordnung*, vol. 8. Heymann.