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

Marc Fabel



| PII: | S0167-6296(21)00016-3 |
|----------------|--|
| DOI: | https://doi.org/10.1016/j.jhealeco.2021.102431 |
| Reference: | JHE 102431 |
| To appear in: | Journal of Health Economics |
| Received Date: | 22 April 2020 |
| Revised Date: | 3 January 2021 |
| Accepted Date: | 6 January 2021 |

Please cite this article as: { doi: https://doi.org/

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

0 Comments

C.1: "1st paragraph, Introduction: According to the data source you provide https://www.oecd.org/gender/data/length-of-maternity-leave-parental-leave-andpaid-father-specific-leave.htm, the average length of paid maternity and parental leave across all OECD countries was 19.08 weeks in 2016 (not 55.2 as you mention). Not sure what the average for these same countries was in 1970, as OECD countries have changed over time, but it seems to be below 16.8. Please revise and choose the same countries over time to make the comparison."

Response:

I am afraid I provided the intermediate link to the database I use. I apologize for the confusion this may have caused. In order to obtain the numbers I cite in the introduction, you have to click on \rightarrow **Download data** at the bottom of the page. This forwards you to the database that provides the desired figures. Since my last visit, the OECD enlarged the set of available countries to 36 member states, which slightly changes the numbers (see Table below). I have adjusted the introduction accordingly (page 1).

| Table 10.1. Deligen of paid materinty and parental leave | Table R.1: | Length | of paid | maternity | and | parental | leave |
|--|------------|--------|---------|-----------|-----|----------|-------|
|--|------------|--------|---------|-----------|-----|----------|-------|

| Year | Average | Median |
|------|---------|--------|
| 1970 | 14.0 | 12.0 |
| 2016 | 55.4 | 45.4 |

Notes: The table shows the average and median length of paid maternity and parental leave (in weeks) for the 36 OECD countries. The indicator is set to 'Total length of paid maternity and parental leave'. Source: OECD, Employment : Length of maternity leave, parental leave, and paid father-specific leave.

https://stats.oecd.org/index.aspx?queryid=54760

The Database reports values for all 36 member states independent of when a country joined the organization. In other words, the averages refer to the same set of countries over time.

To avoid this confusion, I updated the link provided in the list of references (OECD, 2020).

C.2: "Figure 1. I would put "age" in the horizontal axis, rather than year (age). The year(age) labels are confusing, as they are different for treated and control groups."

Response:

Adjusted, as requested (see Figure 1, page 30). Moreover, to have a coherent label throughout the paper, I have also adjusted the horizontal axis in the Figures 2 and 4.

C.3: "Section 4. Empirical Strategy, 2nd paragraph. You claim that seasonality issues are the main reason why a RD deign is not feasible. But seasonality issues would easily be solved by a DiD RD. Your main problem is the imprecision of DiD RD estimates. I would rearrange this paragraph so this is clear (maybe shifting the details about the seasonality discussion to a footnote and shifting the discussion in footnote 28 to the paragraph."

Response:

As suggested, I have rearranged the empirical strategy section (page 9) by shifting the information on the precision problems of discontinuity designs (formerly footnote 28) to the main text and details about season of birth effects to a footnote (footnote 28). When discussing season of birth effects, I explicitly mention that a difference-in-discontinuity design would be a candidate for identification, but must be dismissed in the end due to the lack of precision. I hope the new structure makes everything clearer for the reader.

C.4: "Table 3. You find a positive effect on SST, which is somehow puzzling. Could you provide any intuition for that result? You should at least acknowledge it is unexpected and goes counter to your hypothesis."

Response:

Unfortunately, I cannot provide a satisfactory explanation for the increase in diseases of the skin and subcutaneous tissue. As suggested, I have acknowledged the unexpected increase in footnote 37 (page 13).

C.5: "Reviewer 2's comment about being unable to identify foreigners. You respond that seeking that data would take substantial time. I am fine with it, but you should mention this as a caveat in your discussion or concluding remarks. In any case, it would suggest results are larger than what you are finding due to the dilution effect of migrants."

Response:

To cover the impact of unaffected individuals (e.g. foreigners) on the ITT estimates, I have added footnote 55 (page 22) in the conclusion.

C.6: "Section 6.2 Potential Mechanisms. When addressing gender differences you mention several possible explanations that have to do with parents' differential attention according to gender, but leave aside one explanation you raised before when explaining the results by gender: that men are more likely to suffer from externalizing problem. Parents could have paid the same attention to boys and girls, but girls are more likely to suffer from internalizing problems, and thus may be less likely to end up being hospitalized for that reason (or the larger incidence of depression or anxiety may be seen later on, not as young as the age of 35). I think this latter explanation is more feasible than the former ones you raise in the paragraph."

Response:

I have added your proposed explanation in section 6.2 (page 20). Furthermore, I have deleted the paragraph containing parents' differential attention according to gender.

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

Marc Fabel[†]

January, 2021

Abstract

This paper assesses the impact of the length of maternity leave on children's health outcomes in the long-run. My quasi-experimental design evaluates an expansion in 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. Using this cutoff date as a source of exogenous variation, 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.

Keywords: early childhood development, health, paid maternity leave, long-term child outcomes, life-cycle approach JEL codes: I10, I18, J13, J18, J88

[†]ifo Institute for Economic Research and Munich Graduate School of Economics (email: fabel@ifo.de).

^{*}I am very grateful to Natalia Danzer for valuable comments and discussions. I thank the editor, Ana Balsa, and two anonymous referees who have helped to improve the paper substantially. I also benefited from inspiring feedback from Audrey Au Yong Lyn, Gordon Dahl, Eleonora Guarnieri, Mathias Huebener, Daniel Kühnle, Maarten Lindeboom, David Neumark, Mårten Palme, Erik Plug, Birgitta Rabe, Helmut Rainer, Anna Raute, Patrick Reich, Tanya Wilson and seminar participants at the Annual Meeting of the American Economic Association, the Annual Conference of the European Association of Labour Economists, the IZA World Labor Conference, the Conference of the German Economic Association, the Spring Meeting of Young Economists, and the Ruhr Graduate School Doctoral Conference. Furthermore, I thank Heiko Bergmann for assistance in accessing data. The author gratefully acknowledges financial support from the Leibniz association (P77/2014). Maximilian Grasser and Paulina Hofmann provided excellent research assistance. All errors and omissions are my own.

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 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

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

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 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.

 $^{^{2}}$ 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.

³Enns 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".

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 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

 $^{^{4}}$ 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).

⁵The 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.

⁷Other 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.

⁸The 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).

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 3 explains the data and variables. Section 4 discusses the empirical design. Section 5 reports results. Section 6 contains a discussion of the conceptual framework and mechanisms. Section 7 concludes.

2 Background

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

⁹The 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.

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

¹¹The 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).

¹²Since 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

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.1).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.

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-yearold 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

hours during that year, and if they worked for an employer with at least 50 employees (Baum, 2003).

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

¹⁴Refer 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.

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 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.

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

¹⁷In 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.

¹⁸Both, 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.

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

²⁰Maternal 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.

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.

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.

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 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.1 in the Appendix provides a

²¹Due 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.

²³Please 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.

²⁴The 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.

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.2 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 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 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×months-of-birth×reporting years) observations. In order to abate any confounding effects that might be triggered by 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.

 $^{^{25}\}mathrm{I}$ exclude diagnoses related to "pregnancy, child birth and the puerperium" and others that occur infrequently.

 $^{^{26}}$ 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.

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

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.

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.2 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 differencein-discontinuities design could account for the SOB effects but suffers from the abovementioned 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

$$Y_{mt} = \gamma_0 + \gamma_1 \operatorname{Treat}_m + \gamma_2 \operatorname{After}_m + \gamma_3 (\operatorname{Treat}_m \times \operatorname{After}_m) + \psi_m + \rho_t + \varepsilon_{mt}$$
(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).

²⁹Dustmann 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.

³⁰The 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.

5 Results

5.1 Hospital Admissions

Table 1 reports DiD estimates of the impact of the ML expansion on hospital admissions, using the specification in equation 1. The dependent variable is the hospitalization rate,

³¹The 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 × reduction in maternal labor supply) to obtain the effect of spending one additional month away from work after childbirth on children's health outcomes.

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 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 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 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 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 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 2 display the trajectories of yearly hospitalization differentials over the same time period.³⁵ I estimate the specification in equation 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

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

 $^{^{33}\}mathrm{Yet},$ the baseline means are not significantly different from each other.

 $^{^{34}}$ Table A.4 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.

³⁵I 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.

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.

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 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 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 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

³⁶Panel A of Table A.1 presents an overview of the diagnosis chapters.

³⁷Diseases 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.

³⁸Appendix Figure A.5 shows the respective life-course figures for each diagnosis chapter.

contribution in hospitalization reduction. In the pooled sample, MBDs account for onethird 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; ~ 11 percent of the decline in hospitalizations).

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 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 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 and further increase towards the end of the observed time span. The bottom panels of Figure 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 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

diseases (henceforth called subdiagnoses or subcategories).³⁹ Appendix Figure A.3 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 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.

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 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 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.4 contains a map with the LMR used in the analysis. The dependent variable is using the same denominator

³⁹For 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.1 in the Appendix defines the subcategories of the mental and behavioral diseases chapter.

⁴⁰Appendix Table A.5 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.

⁴¹For 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.

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 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 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:

$$Y_{mt} = \beta_0 + \beta_1 \operatorname{Treat}_m + \beta_2 \operatorname{After}_m + \beta_3 \operatorname{FRG}_m + \beta_4 (\operatorname{Treat}_m \times \operatorname{After}_m) + \beta_5 (\operatorname{Treat}_m \times \operatorname{FRG}_m) + \beta_6 (\operatorname{After}_m \times \operatorname{FRG}_m) + \beta_7 (\operatorname{Treat}_m \times \operatorname{After}_m \times \operatorname{FRG}_m) + \psi_m + \rho_t + \varepsilon_{mt}$$
(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 $\operatorname{Treat}_m \times \operatorname{After}_m$. 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 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

 $^{^{43}}$ The corresponding regression specification is as follows (regions are weighted by population):

 $Y_{mrt} = \gamma_0 + \gamma_1 \operatorname{Treat}_m + \gamma_2 \operatorname{After}_m + \gamma_3 (\operatorname{Treat}_m \times \operatorname{After}_m) + \psi_m + \phi_r + \rho_t + \varepsilon_{mrt}.$

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.7 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 reestimate 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.6 in the Appendix, and show that the main estimates are robust to the different specifications and estimation procedures.

6 Discussion

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 wellbeing 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 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).

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

 $^{^{44}}$ 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).

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

⁴⁶The 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).

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 6. Table 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.

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 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),

⁴⁷Appendix Figure A.4 shows the spatial variation of population density across German labor market regions.

⁴⁸Baker 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.

⁴⁹The 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%.

⁵⁰Higher 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).

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 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

⁵¹See 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.

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 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.

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.

References

- 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. and PISCHKE, J.-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Tech. rep., Princeton University Press.
- AVDIC, D. and KARIMI, A. (2018). Modern Family? Paternity Leave and Marital Stability. American Economic Journal: Applied Economics (4), 283–307.
- BAKER, M. and MILLIGAN, K. (2008). Maternal employment, breastfeeding, and health: Evidence from maternity leave mandates. *Journal of Health Economics* (4), 871–887.
- 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.
- BEUCHERT, L. V., HUMLUM, M. K. and VEJLIN, R. (2016). The length of maternity leave and family health. *Labour Economics*, 55–71.
- 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.
- 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.
- BÜTIKOFER, A., RIISE, J. and SKIRA, M. (forthcoming). The Impact of Paid Maternity Leave on Maternal Health. *American Economic Journal: Economic Policy*.
- CANETTI, L., BACHAR, E., GALILI-WEISSTUB, E., DE-NOUR, A. and SHALEV, A. (1997). Parental bonding and mental health in adolescence. *Adolescence* (126), 381.
- 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/.
- 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.
- 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. 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*.
- 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.

- 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.
- 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.
- FALK, A. and KOSSE, F. (2016). Early Childhood Environment, Breastfeeding and the Formation of Preferences. *HCEO Working Paper* (036).
- 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].
- 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.
- 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.
- HANK, K. and KREYENFELD, M. (2001). Childcare and fertility in (Western) Germany. Max Planck Institute for Demographic Research Working Paper (19).
- 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.
- 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*.

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

- MORRILL, M. S. (2011). The effects of maternal employment on the health of school-age children. *Journal of Health Economics* (2), 240–257.
- 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.
- 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 fatherspecific 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).
- PAPP, L. M. (2014). Longitudinal associations between breastfeeding and observed mother-child interaction qualities in early childhood. *Child: Care, Health and Devel*opment (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.
- 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.
- 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.

- STEARNS, J. (2015). The effects of paid maternity leave: Evidence from Temporary Disability Insurance. Journal of Health Economics, 85–102.
- 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.
- VICTORA, C. G., BAHL, R., BARROS, A. J., FRANÇA, G. V., HORTON, S., KRA-SEVEC, 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.
- ZMARZLIK, J., ZIPPERER, M. and VIETHEN, H. P. (1999). Mutterschutzgesetz, Mutterschaftsleistungen, Bundeserziehungsgeldgesetz: mit Mutterschutzverordnung, vol. 8. Heymann.

FIGURES





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.1.

Figure 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 (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.



Figure 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 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).

Figure 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 (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.



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 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).

TABLES

| | | Estimation window | | | | | | | |
|--|--|--|---|---|--|--|--|--|--|
| | (1) | (2) | (3) | (4) | (5) | | | | |
| | 6M | $5\mathrm{M}$ | $4\mathrm{M}$ | 3M | 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 | | | | |
| $N (MOB \times year)$ | 456 | 380 | 304 | 228 | 380 | | | | |
| | | | | | | | | | |
| Panel B. Age brack | kets | | | | | | | | |
| 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) | | | | |
| Panel B. Age brack Age 17-21 Age 22-26 Age 27-31 Age 32-35 | xets -1.517 (0.946) -0.611 (0.937) -2.665*** (0.826) -3.869*** (1.083) | $\begin{array}{c} -0.590 \\ (0.995) \\ -0.613 \\ (1.113) \\ -3.015^{***} \\ (0.909) \\ -3.619^{**} \\ (1.277) \end{array}$ | -0.735 (1.198) -0.667 (1.412) -3.209^{**} (1.132) -4.572^{***} (1.460) | -1.095 (1.603) -0.735 (1.672) -2.546* (1.375) -5.045** (1.721) | $\begin{array}{c} -1.963^{**}\\ (0.931)\\ -1.080\\ (1.012)\\ -2.974^{***}\\ (0.949)\\ -4.717^{***}\\ (1.191)\end{array}$ | | | | |

 Table 1: ITT effects on hospital admission

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.3 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.

| | | | Women | | | | | Men | | |
|---|---------------|---------------|---------------|---------|----------------|--------------|---------------|---------------|-----------|----------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| | 6M | $5\mathrm{M}$ | $4\mathrm{M}$ | 3M | Donut | 6M | $5\mathrm{M}$ | $4\mathrm{M}$ | 3M | 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 |
| $N (MOB \times year)$ | 456 | 380 | 304 | 228 | 380 | 456 | 380 | 304 | 228 | 380 |
| Panel B. Age brack | cets | | | | | | | | | |
| 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) |

 Table 2: ITT effects on hospital admission, by gender

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.3 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.

| | (1) | (2) | (3) | (4) | (5) |
|----------------|--------------|----------|--------------|--------------|-----------|
| | Overall | | Age brack | kets [years] | |
| | Overall | 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 |

| $\Gamma_{-} [1]_{-} 0.$ | TTTT | - fr f - | | 1 | | 1 | | 1: | -1 |
|-------------------------|------|----------|----|----------|-----------|-----|-----|-----------|----------|
| Ladle 3: | 111 | enects | on | nospital | admission | and | mam | diagnoses | cnapters |
| | | | | | | | | | |

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 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).

| | | Estimation window | | | | | | |
|-----------------------|--------------|-------------------|---------------|---------------|-----------|--|--|--|
| | (1) | (2) | (3) | (4) | (5) | | | |
| | 6M | $5\mathrm{M}$ | $4\mathrm{M}$ | 3M | Donut | | | |
| Panel A. Over ent | ire length o | f the life-co | urse | | | | | |
| 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 brack | kets | | | | | | | |
| 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) | | | |

Table 4: ITT effects on mental & behavioral disorders

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.3 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.

| | | | Women | | | | | Men | | |
|---|-----------|--|--------------|-----------|--------------|----------------|-----------|-----------|---------------|---------------|
| | (1) 6M | $\begin{array}{c} (2) \\ 5 \mathrm{M} \end{array}$ | (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 brack | cets | | | | | | | | | |
| 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) |

Table 5: ITT effects on mental & behavioral disorders, by gender

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.3 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.

| | | Alternative s | specifications | Alter estin | rnative nation | Placebos | |
|--------------------|---------------|-----------------------|-----------------|------------------|-------------------|---------------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | Baseline | current population | LMR $level^a$ | DDD^b | add. CG | temporal: cohort | 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 | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark |
| Year fixed effects | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark |

 Table 6: Robustness checks for hospital admission

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.

| | Heterogeneity | | |
|--|------------------|----------------|--|
| | (1) | (2) | |
| | rural | 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. Montal and behavioral disorders | | | |
| DiD estimate | 0.941 | 0.086*** | |
| DID estimate | (0.564) | (0.300) | |
| Dependent mean | (0.304) 17.00 | 18.61 | |
| Effect in SDs [%] | 1 310 | 7100 | |
| N | 24.987 | 20 568 | |
| 1 | 24,201 | 23,000 | |
| MOB fixed effects | | ✓ | |
| Year fixed effects | • • | • | |
| Region fixed effects | • • | • • | |
| Notes: The table shows DiD estimates of the 1979 n | naternity lea | ave reform on | |

Table 7: Subgroup analysis

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.4 shows a map of Germany with the regions marked as rural/urban.

Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

APPENDIX

(Intended for Online Publication)

V 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.

V.1 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

 $^{^{56}}$ 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.

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. 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 V.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).





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.

 $Source: \mbox{ Birth registry data from North Rhine-Westphalia and Baden-Württemberg. Taken together, both states account for almost 36\% of all births in the former Federal Republic of Germany in 1979.$

Panel B of Figure V.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:

$$Births_i = I_i^{Year} \times I_i^{Day \text{ of Week}} + I_i^{Day \text{ of Year}} + I_i^{Public \text{ Holiday}} + \varepsilon_i$$
(3)

| | | Estimation window | | | | | | | |
|---------------|-----------------------|---------------------|---------------------|---------------------|--|--|--|--|--|
| | (1) | (4) | | | | | | | |
| | $\pm 7~\mathrm{days}$ | $\pm 14~{\rm days}$ | $\pm 21~{\rm days}$ | $\pm 28~{\rm days}$ | | | | | |
| Panel A. Depe | endent vari | able is num | ber 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 | | | | | |

| Table V.1: | Birth | rate | effects | of 1 | the | 1979 | ML | reform |
|------------|-------|------|---------|------|-----|------|----|-------------------------|
|------------|-------|------|---------|------|-----|------|----|-------------------------|

Panel B. Dependent variable is ln(number of births)

| 1 | | (| 0 | / |
|--------------|----------|---------------|----------------|------------|
| 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 \pm 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.

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 (Births – Births) 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 V.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

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

⁵⁹I 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 $I_i^{\text{Day of Week}}$.

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.

V.2 Threats Along the Whole Distribution of Birth Months

As discussed in section 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 2 and 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 V.2 shows estimates with varying bandwidths for the pre-cutoff side by gender and age group. The first estimate per panel, shown in green, represents the baseline estimate as shown in Tables 1 and 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

 $^{^{60}}$ 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.

group born two years before the reform (i.e. born around May 1977, see column 5 of Table 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 V.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 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 V.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 and 2. It uses a bandwidth of half a year.



Figure V.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 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 and 2. It uses a bandwidth of half a year.

APPENDIX FIGURES

Figure A.1: 1979 reform in maternity leave legislation in the Federal Republic of Germany



Length of maternity leave after childbirth

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.2: 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.3: 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.



Figure A.4: 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).





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 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.6: 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 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).



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.

APPENDIX TABLES

| Table A.1: Summary statistic | cs for differ | ent diagnoses | | |
|--|---------------|---------------|---------|--------|
| | (1) | (2) | (3) | (4) |
| | ICD-9 | ICD-10 | Mean | SD |
| | | | | |
| 4. 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 | 800-999 | S00-T98 | 21.196 | 5.978 |
| consequences of external causes | | | | |

 Table A.1: Summary statistics for different diagnoses

Continued on next page

| a | 1 1. 1. | r | 1.00 1 | 1. | / 1 | ۱. |
|-------------|-------------|-----|----------------------|-----------|-----------|----|
| Summarn | STATISTICS | tor | алпетерь | alaanoses | continuea | 1 |
| S anonoar g | 00000000000 | ,0, | <i>auj.) cr cree</i> | aragnooco | contracta | |

| | (1) | (2) | (3) | (4) |
|--|-------------------------------|---------|--------|-------|
| | ICD-9 | ICD-10 | Mean | 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 | 316 | F50-F59 | 0.308 | 0.225 |
| physiological disturbances and physical factors | | | | |
| 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 | $312,\!313,\!314,\!307$ | F90-F98 | 0.320 | 0.535 |
| onset usually occurring in childhood and adolescence | | | | |

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.

| | Es | Estimation window | | | | |
|---------------|-------------|-------------------|---------------|--|--|--|
| | (1) | (2) | (3) | | | |
| | 6M | $5\mathrm{M}$ | $4\mathrm{M}$ | | | |
| Panel A. Over | · entire le | ength of th | e 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) | | | |

Table A.2: RDD on hospital admissions

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.

| | | Estimation window | | | | | | |
|-----------|-----|-------------------|-----|-----|-------|--|--|--|
| | (1) | (2) | (3) | (4) | (5) | | | |
| | 6M | 5M | 4M | 3M | 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 | | | |

Table A.3: Number of observations per age-bracket

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, 2, 4, and 5.

| | | Est | imation win | dow | |
|------------------------------|----------------|---------------|---------------|-----------|----------------|
| | (1) | (2) | (3) | (4) | (5) |
| | 6M | $5\mathrm{M}$ | $4\mathrm{M}$ | 3M | Donut |
| Panel A. Total | | | | | |
| Treatment \times Age 17-21 | -2.310*** | -2.143** | -2.156^{*} | -2.386* | -2.647*** |
| | (0.729) | (0.829) | (1.039) | (1.309) | (0.715) |
| Treatment \times Age 22-26 | -0.0735 | 0.0501 | -0.0501 | -0.00689 | -0.451 |
| | (0.764) | (0.916) | (1.149) | (1.422) | (0.788) |
| Treatment \times Age 27-31 | -2.187** | -1.789 | -2.350 | -2.323 | -2.644** |
| | (0.977) | (1.134) | (1.379) | (1.696) | (1.064) |
| Treatment \times 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 |
| $N (MOB \times year)$ | 456 | 380 | 304 | 228 | 380 |
| | | | | | |
| | | | | | |
| Panel B. Women | | | | | |
| Treatment \times Age 17-21 | -0.879 | -0.607 | 0.153 | -0.292 | -1.370 |
| | (0.992) | (1.048) | (1.176) | (1.560) | (0.946) |
| Treatment \times Age 22-26 | 0.206 | 0.880 | 1.487 | 1.234 | 0.177 |
| | (0.943) | (1.058) | (1.278) | (1.698) | (0.872) |
| Treatment \times Age 27-31 | -3.083*** | -2.204* | -1.923 | -2.429 | -3.404*** |
| | (1.060) | (1.119) | (1.338) | (1.790) | (1.035) |
| Treatment \times 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 |
| $N (MOB \times year)$ | 456 | 380 | 304 | 228 | 380 |
| | | | | | |
| | | | | | |
| Panel C. Men | | | | | |
| Treatment \times Age 17-21 | -3.738*** | -3.635*** | -4.360*** | -4.397** | -3.937*** |
| | (0.984) | (1.110) | (1.317) | (1.561) | (1.133) |
| Treatment \times Age 22-26 | -0.352 | -0.751 | -1.513 | -1.193 | -1.060 |
| | (1.072) | (1.264) | (1.521) | (1.696) | (1.212) |
| Treatment \times Age 27-31 | -1.325 | -1.402 | -2.759 | -2.224 | -1.908 |
| | (1.318) | (1.585) | (1.811) | (1.914) | (1.548) |
| Treatment \times 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 |
| $N (MOB \times year)$ | 456 | 380 | 304 | 228 | 380 |

Table A.4: Robustness: Interaction Treatment with Age Brackets (hospitalization)

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 in which there are different regressions for each age-bracket.

| | (1) | (2) | (3) | (4) | (5) | | |
|-------------------------|-----------|----------------------|---------|----------------|-------------|--|--|
| | Overall | Age brackets [years] | | | | | |
| | Overair | 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 | | |

 Table A.5: ITT effects on the subcategories of mental and behavioral disorders

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 for additional details. Clustered standard errors are reported in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

| | | Alternative specifications | | Alter estim | Alternative estimation | | Placebos | |
|--------------------|--------------|----------------------------|---------------------------|------------------|---------------------------|---------------------|-----------------|--|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | |
| | Baseline | current population | LMR level ^a | DDD^b | add. CG | temporal: cohort | 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 | |
| Ν | 456 | 288 | $53,\!855$ | 912 | 720 | 456 | 456 | |
| MOB fixed effects | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | |
| Year fixed effects | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | \checkmark | |

 Table A.6: Robustness checks for mental and behavioral disorders

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.