2012-6
Miriam Wüst
PhD Thesis

Essays on early investments in child health

DEPARTMENT OF ECONOMICS AND BUSINESS
AARHUS UNIVERSITY • DENMARK

PhD-Thesis

Essays on early investments in child health

Author:
Miriam Wüst

Supervisors:
Nabanita Datta Gupta
Tor Eriksson
Paul Bingley

May 2, 2012

Department of Economics and Business
Aarhus University, Hermodsvej 22, 8230 Aabyhøj, Denmark,
and
SFI-The Danish National Centre for Social Research
mirw@asb.dk, miw@sfi.dk

Preface

This thesis is the result of my PhD studies at the Department of Economics and Business, Aarhus University (AU) and SFI-The Danish National Centre for Social Research. I benefited greatly from the support of both institutions, which provided me with the financial means to travel to conferences and courses, with data for my projects and with great workplaces. I am indebted to many people who have made the past three years an instructive time.

First and foremost, I thank my supervisors Nabanita Datta Gupta, Tor Eriksson and Paul Bingley who have been extraordinarily committed and have supported me greatly with their invaluable advice and guidance. Their ever positive and constructive comments have improved my work and encouraged me to go the extra mile.

I also thank my PhD committee—Anna Piil Damm (AU), Christina Gathmann (University of Heidelberg) and Maarten Lindeboom (Vrije University)—for thorough reading, very helpful comments, and an in-depth discussion of all chapters during my pre-defense. While some comments have already resulted in changes to this final version of the thesis, I will incorporate all their suggestions in future work on the papers.

I thank my colleagues at SFI and AU—importantly Lena Janys (for shelter in Aarhus and great company), Vibeke Myrup Jensen (for a latte when I needed it most), my fellow PhD students at both institutions, my office mates at "the 4^{th} floor" and "the office 301", and the members of the SFI Health Research Group—all of you inspired me and made my PhD studies an enjoyable time. A special thanks goes to Jon Kvist, who inspired me to work in research in the first place and told me to read more. Real good advice!

From February 2011 to May 2011 I visited the Department of Economics at Columbia University, NYC. I thank Professor Janet Currie for inviting me and taking time to discuss my work, and fellow graduate students for making my stay a productive and inspiring time. For their generous financial support of my visit, I thank the Aarhus University Research Foundation, the Otto Mønsted Foundation and the Oticon Foundation. Thanks also to my former Head of Department Peter Jensen, AU for his commitment to securing the funding for my visit.

I am indebted to a number of unselfish colleagues, who shared their data with me. I thank Anne-Marie Nybo Andersen who kindly provided me with data and advice for the first chapter of this thesis, and Kirsten Krøyer and the DigDag-project, who helped me with my treasure hunt for historical data for the second chapter. I also thank my SFI colleague Mette Lausten, who made the data for the third chapter available.

Furthermore, I was lucky to work with Natalie Reid while writing this thesis. Thanks for teaching me invaluable lessons on academic writing. Susan Stilling and Camilla Mikkelsen made all administrative matters appear small and easy to solve, and I am grateful for their help. In the final stressful days, Susan Stilling and Signe Hastrup Poulsen provided invaluable help with proofreading and making my Danish summary worth reading!

A huge Danke goes to my family, my parents Annemarie and Fred Wüst, and my brother Alexander, for their unconditional love and support of all my endeavours. I think of you every day. Finally, I thank you, Morten, for being there in good times and in bad—just like you promised. I am looking forward to all the good times ahead of us and hope your promise extends to the postdoc, too!

Miriam Wüst
Copenhagen, March 2012, updated May 2012

Contents

Introduction

This dissertation contains three independent essays which evolve around the question on the ways in which parental and public investments impact birth outcomes and infant health. My analysis is motivated twofold. First, children do not choose their parents. The "accident of birth" (Cunha and Heckman, 2007) assigns them to their families. Given that promoting health equality at birth and early in life is a central goal of public health policy, knowledge on the importance of various health inputs is instrumental to policy makers.

Second, a growing economics literature—which complements a similar literature in epidemiology—has examined the ways in which health at birth and in infancy affects adult outcomes such as education, labor market participation, and health (Behrman and Rosenzweig, 2004; Case et al., 2005; Black et al., 2007; Currie, 2009). This literature has documented a strong link between conditions in utero, health at birth and in infancy, and later outcomes and—importantly—that this link is a causal one (for an overview see Almond and Currie, 2011). Consequently, policies that promote early life health potentially have high returns over and above the immediate ones.

The empirical literature on "early influences" has also stimulated economists to reconsider the ways in which we think about and model health and human capital formation. While the classical Grossman model has been at the core of economists' theoretical understanding of health production (Grossman, 1972), it has important constraints when applying it to the analysis of early investments in health. The empirical finding that early health investments and shocks matter for outcomes in adulthood—i.e. that the impact of early life events does not necessarily depreciate over time—has inspired new theoretical work which adds to the basic Grossman model, among other things with a focus on timing of investments (see among many influential contributions Cunha et al., 2006).

The timing of parental and public investments is of major importance in these models of skill formation for at least three reasons: first, there are critical periods of investment, i.e., certain investments are specific to or at least most efficient in certain periods. Pregnancy is likely to be a critical period that sets the stage for later investments. Thus a better understanding of parental choices and the effects of prenatal investments is highly relevant. Second, the concept of "dynamic complementaries" suggests that certain early investments are hard to substitute with later ones, i.e. the model relaxes the assumption that inputs in different childhood periods are perfect substitutes. Third, early investments have the potential to operate through many childhood periods and are thus likely to be pivotal, or as stated in the concept of "self productivity", early acquired skills feed through many periods and generate later skills and abilities. Given that certain early investments in the health of infants are hard to substitute and have long-lasting consequences, *identifying* those investments and promoting them by means of policy has the potential for important individual and societal returns.

This thesis empirically investigates three different aspects of parental and public investments all of which aim at improving infants' health. The first chapter investigate the role of prenatal maternal health behaviors, most importantly maternal prenatal employment. As stated by the fetal origins hypothesis and shown in many empirical studies, conditions in utero are decisive for various later life outcomes (for an overview see Almond and Currie, 2011a). This chapter examines the ways in which maternal prenatal behaviors modify birth outcomes.

The second chapter examines the effects of a universal home visiting program for infants and their mothers on the infant mortality rates of Denmark in the 1930s and 1940s. Importantly, this chapter examines a policy and the treatment is clearly defined in content and duration. The analysis of the causes of the infant mortality decline in the early 20^{th} century is informative in its own right. Furthermore, exposure to high rates of infant mortality has been shown to predict future adverse outcomes (see e.g. Almond, 2006).

While the chapter's main focus is on immediate health effects, it also takes a first glance at longer-run returns to the program. Thus the chapter locates itself in a recent and vivid literature on early influences that has exploited historical data on natural experiments—among them pandemics, wars, and medical and public health advances (Cutler and Miller, 2005; Almond and Chay, 2006; Almond, 2006; van den Berg et al., 2006; Bhalotra and Venkataramani, 2011). Economists have contributed to this literature by suggesting methods to establish a causal link of conditions in infancy and adult outcomes, by finding and generating new data, often based on aggregated historical records, and by broadening the focus of epidemiologists on adult health to other outcomes such as educational attainment, wages, fertility, mariage behavior and cognitive functioning in adulthood.

The third chapter investigates the health effects of Caesarean section (CS) for childbirth. Economists' interest in CS use has earlier almost exclusively been motivated by the costs induced to the medical system. In contrast, we examine the costs and benefits of CS use for the patient, in our study the specific group of pregnancies with babies in breech presentation. Furthermore, our study investigates the impact of newly available scientific information on procedure use, i.e. on the behavior on the supply side of the medical system.

The chapters of this thesis investigate the short-run effects of parental and public investments. By analyzing immediate outcomes—birth weight and premature birth, and infant morbidity and mortality—I zoom in on outcomes that themselves constitute some of the determinants of subsequent investments and individuals' future circumstances.

While short-run outcomes are at the center of this thesis, the chapters form the foundation for future research projects on longer-run consequences. A natural extension to the chapter on home visiting, for example, is to incorporate better data and additional adult-life outcomes into the analysis of the effects of the program, a public investment already made in individuals' infancy. Thus the present analysis of the determinants of early life health lays the first stones for my future analyses of long-run returns to early investments in infants' health.

Summary of the chapters

In this section I briefly summarize the three chapters, their data and methods, their main findings and implications. While focusing on the same overarching theme of early health investments, the three chapters of this thesis capitalize on data from different sources and adapt to the threats to identification in their specific ways.

Chapter one is entitled *"Maternal Employment during Pregnancy and Birth Outcomes"*. While economic research on prenatal health inputs and their effect on birth outcomes has focused on several modifiable behaviors—among them maternal smoking, prenatal care initiation and choices made long before pregnancy, such as maternal education—this chapter contributes to the literature by focusing on a until recently neglected maternal input, namely maternal employment during pregnancy. Given rising female employment rates in most developed countries, a better understanding of the effects of maternal work during pregnancy on birth outcomes is warranted and of great policy importance.

I use high-quality register and survey data that combines the best of two worlds: first, high-quality outcome measures and, second, information on behaviors during pregnancy that was collected contemporarily (i.e., during pregnancy). Exploiting variation between siblings to account for the endogeneity of prenatal input decisions, my findings illustrate that universal claims about the harmful effect of maternal employment on birth outcomes have no bearing. Rather than that, conflicting results to findings from the U.S. and the UK illustrate that maternal employment has to be studied in its specific context. The analysis for Denmark—a country with high female employment rates also late during pregnancy and generous parental leave—exemplifies that maternal employment in the third trimester does not inevitably lead to worse birth outcomes. Thus policies that further extend prenatal leave to above one to two months before due date may not be effective in improving birth outcomes.

I find that mothers who work in the last trimester (around week 30 of their pregnancy) have a lower probability of premature birth. While I do not find strong results for birth weight, my results for premature birth are robust across specifications. Furthermore, they indicate that blue-collar occupation mothers even benefit more. As this finding is not driven by selective attrition from the labor market according to occupation, it indicates that the adaptation of work tasks for pregnant women (granted by law) and the expectation of a generous maternity leave take pressures from mothers' shoulders. This mechanism could be of even greater importance for blue-collar mothers who face more manual labor in general and for whom adaptation of work tasks assumingly is largest.

Another informative finding from this chapter is that a high socio-economic status not automatically implies favorable health behaviors—such as abstaining from alcohol consumption—on the part of the mother. This finding suggests that common claims about the positive correlation of a high socio-economic status and health behaviors during pregnancy are not valid per default. Thus strategies to encourage positive health behaviors during pregnancy should be designed to take the heterogeneity of maternal behaviors into account.

Chapter one also points at the direction for future research: little is known as to date about the nature of parental response to child endowments. This lack of knowledge is a challenge for estimates based on sibling comparisons in this literature. Thus future studies should examine whether poor birth outcomes stimulate compensating or reinforcing parental investments.

Chapter two turns to the analysis of a policy response designed to improve infants' health and its effects. The chapter *"Early Interventions and Health Outcomes: Evidence from the Danish Home Visiting Program"* investigates the introduction of universal home visiting and its effect on infant survival rates in Denmark in the 1930s and the 1940s. The paper is based on unique historical data from archives, national reports and registers, all of which I combined for this purpose.

I exploit exogenous variation in the timing of the introduction of the treatment across municipalities. This variation results from the decentralised implementation of the program and educational constraints which created shortages of nurses. I find that the home visiting program significantly contributed to the overall decrease in infant mortality in the period considered and it did so at relatively modest costs. My results are robust across specifications and live up to several robustness tests.

The visiting nurses heavily promoted maternal breastfeeding and a proper home environment. Thus the effect of the program was most likely through two channels, improvements of infant nutrition and hygienic conditions, and timely referral of ill children to doctors. In the period considered, infectious diseases—many of which resulted from poor hygienic conditions—were still a major killer. The same situation still holds for many developing countries today and recent research shows that many deaths could be prevented at low costs with programs that bring health professionals to families at an early time. Universal home visiting could potentially play a role in this context.

The early influence literature suggests that early life events can have persistent effects that do not necessarily fade out. Thus early interventions—like home visiting—have the potential for important long-run effects. Consequently, the chapter takes a first glance at long-run returns to the program, and I find indication for treated individuals being shorter and more likely to be overweight at their military examination. This finding is consistent with a scarring story, where weaker individuals survive in treated areas and have poorer long-run outcomes. However, as this finding is based on a small sample of twins, it should be interpreted very cautiously. My future work on the effects of universal home visiting will pursue this journey further by connecting the collected information on the expansion of the nurse program to population data from the Danish registers. This step will open new possibilities for examining the returns to this Danish public health initiative.

The third and final chapter is related to the contemporary scientific and public debate on medical procedure use for childbirth. It is entitled *"The Effect of Caesarean Section for Babies in Breech Presentation on Child and Mother Health. Evidence from a Regression*

Discontinuity Design". The chapter is joint work with Vibeke Myrup Jensen, SFI.

Specifically, this chapter uses Danish administrative register data for all breech babies born between 1997 and 2006. We exploit a discontinuous change in CS probability for breech babies around the dissemination of the results of a famous Lancet study in 2000 (the "Term Breech Trial", TBT). The study—which was based on randomization—found that planned CS is superior to planned vaginal birth for babies in breech position that met the inclusion criteria of the trial. While the study was very influential, its conclusions and recommendations have been criticized on many grounds ever since.

In contrast to the *Intention to Treat* analysis of the TBT, we perform a *Treatment on the Treated* analysis. We use a fuzzy regression discontinuity design. In a policy field where randomization has little public support, we suggest that the use of designs like the regression discontinuity approach can be fruitful alternatives which we can use to identify local treatment effects.

We find that performed CS results in a lower probability of having a low APGAR score at five minutes—which is a global indicator of newborn health—for the marginal child. We furthermore find some indication for CS use decreasing the probability of having an above average number of visits at the general practitioner in the first two years of life.

We find no significant results for other measures of health care utilization or mother health outcomes directly after birth. We conclude that CS use has no detectable negative effects on the outcomes we consider for the marginal breech baby and her mother. Future work will follow up on the potential longer-run benefits of CS use.

References

Almond, D, "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy*, 2006, *114*, 672–712.

_ and J Currie, "Chapter 15: Human Capital Development before Age Five," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4, Part 2 of *Handbook of Labor Economics*, Elsevier, 2011, pp. 1315–1486.

_ and _ , "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, 2011a, *25* (3), 153–72.

_ and KY Chay, "The Long-Run and Intergenerational Impact of Poor Infant Health: Evidence from Cohorts Born During the Civil Rights Era," *Working paper*, 2006.

Behrman, JR and MR Rosenzweig, "Returns to Birthweight," *The Review of Economics and Statistics*, 06 2004, *86* (2), 586–601.

Bhalotra, S and A Venkataramani, "The captain of the men of death and his shadow: Long-run impacts of early life pneumonia exposure," IZA Working Paper 2011.

Black, SE, PJ Devereux, and KG Salvanes, "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *The Quarterly Journal of Economics*, 02 2007, *122* (1), 409–439.

Case, A, A Fertig, and C Paxson, "The lasting impact of childhood health and circumstance," *Journal of Health Economics*, March 2005, *24* (2), 365–389.

Cunha, F and J Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

_ , _ , L Lochner, and DV Masterov, "Chapter 12 Interpreting the Evidence on Life Cycle Skill Formation," in E. Hanushek and F. Welch, eds., *E. Hanushek and F. Welch, eds.*, Vol. 1 of *Handbook of the Economics of Education*, Elsevier, 2006, pp. 697 – 812.

Currie, J, "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development," *Journal of Economic Literature*, 2009, *47* (1), 87–122.

Cutler, D and G Miller, "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States.," *Demography*, 2005, *42* (1), 1–22.

Grossman, M, "On the Concept of Health Capital and the Demand for Health," *The Journal of Political Economy*, 1972, *80* (2), 223–255.

van den Berg, G, M Lindeboom, and F Portrait, "Economic conditions early in life and individual mortality," *American Economic Review*, 2006, *96*, 290–302.

Danish Summary

Denne afhandling består af tre uafhængige kapitler. Selvom de tre kapitler bruger forskellige datakilder og forskellige økonometriske metoder, er det overordnede mål det samme: at afdække kausale effekter af tidlige investeringer i børns sundhed.

Der er mindst to grunde til, at økonomer interesserer sig for disse tidlige investeringer. For det første viser en stor og voksende forskning inden for økonomi og epidemiologi, at sundhed tidligt i livet er en god prædiktor for, hvordan børn klarer sig senere i livet. Der er altså grund til at antage, at indsatser, som forbedrer småbørns sundhed, giver afkast mange år senere. For det andet vælger vi ikke selv vores forældre. Hvis vi mener, at samfundet har et ansvar for at sikre lige muligheder for alle børn, så er indsatser, som sikrer, at alle børn får en sund start på livet, særligt vigtige. For at kunne udvælge den bedste indsats blandt mange mulige skal vi vide, hvilke indsatser der virker. Denne afhandling bidrager til at øge denne viden.

Det første kapitel, "Maternal Employment during Pregnancy and Birth Outcomes", handler om årsager, der kan føre til lav fødselsvægt og for tidlig fødsel. Jeg undersøger, hvilken rolle morens arbejdsmarkedsdeltagelse spiller for barnets sundhedstilstand ved fødslen. I mange industrilande er andelen af kvinder, som arbejder under graviditeten, stigende, men det er ikke undersøgt særlig godt, hvilke konsekvenser denne udvikling har for børns sundhed ved fødslen. Viden inden for dette område er vigtig, når fx barselsordninger skal designes.

Jeg bruger danske survey- og registerdata for at identificere betydningen af morens arbejdsmarkedsdeltagelse under graviditeten. De danske data er særlig velegnede til dette formål: De indeholder mange oplysninger om morens karakteristika, de er målt med meget lidt målefejl, og de indeholder oplysninger om søskende og kan derfor bruges til at eliminere den observerede heterogenitet mellem mødre, som ofte volder problemer i den slags analyser.

I en dansk kontekst - med mange kvinder på arbejdsmarkedet og generøse barselsordninger - har morens arbejde under graviditeten ifølge mine analyser ikke nogen negative konsekvenser for hverken barnets fødselsvægt eller sandsynligheden for, at moren føder for tidligt. I modsætning til tidligere resultater for USA og Storbritannien finder jeg, at kvinder, som arbejder, har en lavere sandsynlighed for at føde for tidligt, mens barnets fødselsvægt er upåvirket. Jeg finder desuden større effekter for kvinder med manuelt præget arbejde. Dette kunne skyldes, at deres arbejdsgivere er bedre til at tilrettelægge arbejdsopgaver, eller at de profiterer af at være udsat for kollegaer (såkaldte "peer effects").

Det andet kapitel, "Early Intervensions and Health Outcomes: Evidence from the Danish Home Visiting Program", belyser effekterne af etableringen af den danske sundhedsplejerske-ordning i 1937. Jeg undersøger, om ordningen har bidraget til faldet af spædbørnsdødelighed i perioden op til 1949. Desuden indeholder kapitlet en første analyse af ordningens langtid-seffekter.

For at undersøge effekterne af sundhedsplejerskeordningen bruger jeg historiske data for alle danske købstæder og for de danske amtslægekredse. Desuden har jeg indhentet data fra Rigsarkivet om, hvornår de forskellige kommuner har indført den frivillige sundhedsplejerske-ordning. Jeg bruger denne information til at identificere effekten af ordningen. Jeg finder, at den har bidraget til faldet af spædbørnsdødelighed i perioden. Jeg tester dette resultat på forskellige måder, blandt andet ved at introducere forskellige trends og kontrolvariable i mine analyser. Desuden finder jeg, at sundhedsplejerskeordningen særligt har bidaget til fald i dødligheden som følge af kolerine – tidligere en af de hyppige dødsårsager blandt spædbørn, som ofte skyldtes dårlig madhygiejne. Sundhedsplejerskerne promoverede amning, og at lige præcis dødligheden for kolerine faldt så kraftigt, som den gjorde, må ses som et udtryk for deres succes.

Eftersom den økonomiske teori forudsiger, at hændelser tidligt i livet kan have betydning for udfald senere i livet, er det interessant at undersøge, om man kan spore effekterne af sund-hedsplejerskeordningen i voksenalderen. Mine foreløbige analyser er baseret på det danske Tvillingeregister, og de antyder, at tvillingedrenge som blev født i kommuner med en sund-hedsplejerske, var lavere og havde større sandsynlighed for at være overvægtige ved sessionen. Dette resultat kan muligvis skyldes, at flere svage børn overlevede i indsatskommuner. Det er vigtigt at huske at dette foreløbige resultat er baseret på en ikke-repræsentativ stikprøve. I fremtidige analyser vil jeg bruge danske registerdata for hele populationen (årgangene født mellem 1933 og 1949) for at undersøge langtidseffekter af sundhedsplejerskeordningen.

Det tredje kapitel, "The Effect of Caesarean Section for Babies in Breech Presentation on Child and Mother Health. Evidence from a Regression Discontinuity Design'", er blevet til i et samarbejde med Vibeke Myrup Jensen (SFI). Vi undersøger, hvilke konsekvenser en fødsel ved kejsersnit har for børns og mødres helbred. Vi ser på en særlig gruppe af fødsler, nemlig sædefødsler. I 2000 publicerede The Lancet et betydningsfuldt studie (The Term Breech Trial), som viste, at kejsersnit er bedre for denne gruppe børn end en naturlig fødsel. Som konsekvens af denne undersøgelse ændrede danske læger deres praksis i forhold til denne gruppe, dvs. langt flere børn i sædestilling blev født ved kejsersnit, efter studiet blev offentliggjort og diskuteret i Danmark.

Vi bruger den abrupte ændring i lægers adfærd til at bestemme effekten af kejsersnit i et "fuzzy regression discontinuity design". Vi finder, at sandsynligheden for kejsersnit stiger kraftigt for den udvalgte gruppe af flergangsfødende med barn i sædestilling. Desuden finder vi indikationer på, at det marginale barn født ved kejsersnit (på grund af TBT publiceringen) har en lavere sandsynlighed for at have en lav APGAR-score fem minutter efter fødslen. Det marginale barn har også en lavere sandsynlighed for at have mange besøg hos almen læge i de første to år. For andre helbredsmål – blandt andet antal indlæggelser og ambulante sygehusbesøg – finder vi ingen effekter. Det samme gælder for efterfødselskomplikationer for mødrene.

Maternal employment during pregnancy and birth outcomes

Miriam Wüst[*]

Abstract

This paper uses high quality Danish survey and register data to analyze the effect of maternal prenatal inputs on birth weight, fetal growth, and the likelihood of premature birth. Given rising female employment rates, the paper focuses on the effect of maternal employment during pregnancy net of other maternal prenatal health inputs. To account for the endogeneity of maternal employment, I exploit information on children who are siblings. This paper is the first to incorporate a wide array of control variables—including maternal occupational status prior to pregnancy, pregnancy-related health problems and sick listing from the job, and a set of maternal health behaviors such as smoking and alcohol consumption during pregnancy—into the sibling analysis. I consistently find that maternal work during the last trimester of pregnancy reduces the risk of premature birth. For birth weight and fetal growth, I find no significant employment effect. As my findings are in contrast to results for the UK and the U.S., they highlight the importance of country-specific factors such as maternity leave. Furthermore, the findings suggest that stress on the part of the mother could be due to exclusion from the labor market rather than work. Finally, I find negative smoking effects on birth weight of conventional size, and negative effects for alcohol consumption on all outcomes. The alcohol effect is driven by well-educated mothers and thus calls into question common assumptions about the positive correlation of favorable health behaviors with proxies for high socio-economic status.

[*]I thank Nabanita Datta Gupta, Tor Eriksson, Paul Bingley, Janet Currie, Maya Rossin-Slater, Petter Lundborg, Anne-Marie Nybo Andersen, participants at the 2009 ESPE conference, and seminar participants at SFI and AU for helpful comments.
I acknowledge financial support by the Danish Agency for Science, Technology and Innovation through a grant to the Graduate School for Integration, Production and Welfare. This paper uses data from the Danish National Birth Cohort (DNBC), which was created by the Danish Epidemiology Science Centre. The cohort was funded by the Danish National Research Foundation. Additional support was obtained from the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation. I thank Anne-Marie Nybo Andersen for providing me with the data.

1 Introduction

Recent economic research suggests the existence of sensitive and critical periods for investment in children, and pregnancy is likely to be among the critical ones (Almond and Currie, 2011; Cunha and Heckman, 2007). Furthermore, there is growing evidence for substantial returns to birth outcomes for long-run outcomes such as educational attainment, labor market attachment, and health (Oreopoulos et al., 2008; Black et al., 2007; Currie et al., 2010). Thus a better understanding of the effect of prenatal inputs on child health at birth is needed.

In most developed countries, the percentage of employed pregnant women and mothers has steadily increased over the last three decades. At the same time, first-time mothers are older and have higher educational attainment. Nevertheless, while a number of studies have focused on the effect of maternal age and education on child health at birth, the role of mothers' prenatal employment is under-researched. In the U.S., for example, around 67 percent of first time mothers worked during their pregnancy in the period 2001-2003 (Johnson, 2007)—87 percent of whom worked during the last trimester and 64 percent of whom worked up to the ninth month. Given that an increased percentage of mothers work during pregnancy, knowledge on the effects of prenatal employment on birth outcomes is instrumental for policy makers when designing policies such as maternity leave.

As for other prenatal inputs, recovering the effect of maternal employment is complicated by omitted variables: Employment during pregnancy is not randomly distributed among women but dependent on various characteristics, only some of which are observable to the researcher. Among those differences are women's job and workplace characteristics, their health status, their preferences and financial constraints, and their access to maternity leave. Furthermore, if women's employment decisions are correlated to other time-varying behaviors such as smoking, alcohol consumption, or the initiation of prenatal care, and if these behaviors have an impact on birth outcomes, estimates of the employment effect will be biased in the absence of adequate control.

One recent study by economists explicitly incorporates maternal employment—i.e., actually working during the last trimester as opposed to just formally holding a job and *not* working—in its analysis of the determinants of birth outcomes in the UK and the U.S. Using data on siblings, and thereby taking mother-specific unobservables into account, Del Bono et al. (2008) find that not working or taking work interruptions up to two months before birth has positive effects on birth weight and fetal growth. Moreover, in their British sample, they find that mothers with low education benefit more from work interruptions during pregnancy than mothers with high education. However, 60 percent of the mothers in their British sample are either non-working during pregnancy or there is no information on their employment status. This figure indicates that improved birth outcomes reflect the benefit of maternity leave for financially constrained women who remain employed during pregnancy. Similarly financially constrained mothers in the U.S. most likely cannot afford to take maternity leave and be absent from work.

In a related study on the impact of maternity leave, Rossin (2011) finds that the enactment of unpaid maternity leave in the U.S. improved child health at birth and infancy for college-educated and married mothers. She finds that while disadvantaged mothers cannot afford to take maternity leave, mothers who are likely to take advantage of the unpaid prenatal leave and take work interruptions before birth experience improvements in their children's birth weight and decreases in the probability of a premature birth (Rossin, 2011). This finding is consistent to the findings in Del Bono et al. (2008). Unfortunately, Rossin (2011) cannot link mothers' employment status during pregnancy or information on their occupational status to birth outcomes. Still, both studies point at the importance of maternity leave and its design for birth outcomes.

Virtually no other study in economics has studied the role of prenatal maternal employment on birth outcomes. Yet, because the impact of maternal employment is likely to vary across countries, and because the existing studies face several data constraints, a number of questions remain unresolved. This paper examines three of these issues. First, by using data from Denmark, this paper examines the effect of maternal prenatal employment in an institutional setting with both high female labor market participation rates and generous maternity leave. I study the effect of actually being at work during the last trimester as opposed to only formally holding at job.

Given that I am interested in measuring the effect of being at the workplace, and as pregnant Danish women have options to leave the labor market in a number of ways, I put emphasis on measuring women's employment status accurately. Pregnant women in Denmark can leave the labor market on compensated prenatal leave at around four weeks before their due date. At the same time, most Danish women work up to this prenatal leave.[1] While pregnant women usually continue in their pre-pregnancy jobs, they are granted by law that work tasks are adapted to their pregnancy-related needs and they can leave their job on compensated sick leave if necessary. These conditions suggest that women with physically demanding jobs face different circumstances during their pregnancy than women in other countries without similar workplace adjustments, i.e. they are potentially more likely to stay on the labor market because of workplace adjustments.

At the same time, another institutional feature in Denmark allows mothers with health problems to leave the labor market on compensated sick leave. This sick leave may be used more extensively by mothers with physically demanding jobs. Thus controlling for maternal occupation, maternal health and whether the mother actually is at work—or on sick leave—is important to disentangle the different potential mechanisms for an effect of maternal employment on birth outcomes.

As a second contribution, the paper factors in detailed information on other prenatal maternal health inputs from high quality survey and register data. From the national registers I use information on maternal occupational status in the year before birth, maternal age and education, and maternal diagnoses for pregnancy-related health problems. The survey holds information on a number of relevant health behaviors such as maternal work and sick leave from work during pregnancy, maternal pre-pregnancy body mass index (BMI), maternal smoking, alcohol and fish consumption (the latter as a proxy for healthy diet), and maternal exercise during pregnancy. Because mothers responded to the survey during pregnancy and the register data on inputs and outcomes is linked to the survey by a unique personal identifier, I do not encounter recall or matching problems in my analysis. As a third contribution, I consider an additional birth outcome that Del Bono et al. (2008) do not consider but that studies using cross-sectional data have shown to correlate with maternal employment: the probability of premature birth.

Using a sibling comparison, I find a protective role for maternal employment in the third trimester with respect to the likelihood of premature birth. Women who work late in their pregnancy (measured at around 8.5 weeks before their due date) are 2.8 percentage points less likely to experience a premature birth. This result is robust to a number of informal tests and

[1] Women's prenatal leave is regulated in collective agreements and some groups have up to 8 weeks prenatal leave. I measure maternal work around week 30, i.e. before women are on prenatal leave.

different specifications. At the sibling sample mean of 3 percent, the effect of maternal work is of considerable size. My results for the effect of maternal employment on birth weight—in contrast to Del Bono et al. (2008)—are also positive but very small and imprecisely estimated. Thus I do not find important effects for birth weight. As this measure is much more affected by conditions late in pregnancy (where the baby grows rapidly), this finding could reflect the effect of a generous prenatal maternal leave for almost all women irrespectively of whether they have worked during pregnancy.

My findings could result from alternative and not mutually exclusive underlying mechanisms. Given generous welfare benefits and universal and free prenatal care, I judge the income channel—working mothers have a better income—to be of minor importance for the effect of prenatal employment on birth outcomes. Rather than that, another mechanism frequently suggested in the literature could be at work: My findings suggest that the stress mechanism potentially works in the opposite way in Denmark as earlier found, namely by inducing stress to those mothers who do not work during their pregnancy. Those mothers could be more worried about their future labor market prospects after birth. Another potential mechanism for the positive employment effect is the presence of peers and networks at the workplace. Recent research—from Scandinavia—suggests peer effects in the timing of pregnancy for women at the same workplace (Hensvik and Nilsson, 2010; Ciliberto et al., 2010). These effects could translate in better birth outcomes either through better support for pregnant women, peer effects that impact women's behaviors during pregnancy, or more optimal timing of fertility (e.g. because mothers delay pregnancy).

Estimating the employment effect separately for white- and blue-collar mothers reveals that the protective employment effect with respect to premature birth is more pronounced in the blue-collar occupational groups (i.e., ISCO groups five to nine). This result could both be driven by better workplace accommodation for this group (if employers are more aware of the need to adapt work tasks for these mothers) or by doctors being more prone to sick list mothers in the blue-collar group. If blue-collar mothers are on sick leave at a lower health threshold than white-collar mothers, the bigger effect in the blue-collar group could be partly driven by blue-collar mothers leaving the labor market more often (which would make the employed blue-collar mothers a select group of very healthy women). A final possibility is that blue-collar mothers experience even more stress from not being employed, e.g. because they assume that it is harder for them to find a job again after birth.

My findings casts doubt on common arguments of stress exposure and its harmful effects for women who work during the last trimester of pregnancy. Given that Danish workplaces are obligated to accommodate to pregnant women's needs, that women can obtain periods of paid sick leave during pregnancy, and that close to all women have access to paid maternity leave one month before the due date, the results indicate that these institutional features contribute to different conclusions about the health effects of prenatal employment in Denmark than those for the U.S. and the UK.

In addition to my results for prenatal maternal employment, I find effects of conventional size for maternal smoking on birth weight. Smoking reduces birth weight by around 94.5g and this effect is driven by mothers in the low-education (i.e., mothers who only completed high school) and blue-collar occupational groups. As this finding is in line with earlier results based on sibling comparison, it suggests that those studies—despite their lack of detailed control for other health inputs—do well in estimating the causal effect of prenatal smoking on birth outcomes.

Finally, I find a negative effect of maternal alcohol consumption on the risk of premature birth and birth weight. At the mean of around two weekly units (for those who have a non-zero consumption), this negative effect amounts to an increase of 1.6 percentage points for the likelihood of premature birth and a very small decrease of birth weight by around 36g. The results suggest a strong dose-response relationship. As the negative alcohol effect is driven by mothers with high education and mothers in the white-collar occupational groups, this pattern contradicts studies that, for lack of data, assume a correlation of maternal education with positive health behaviors during pregnancy (Abrevaya, 2006). For Denmark, this finding is in line with earlier studies showing that while high-education mothers abstain from binge drinking during pregnancy, they find a moderate consumption of alcohol acceptable (Strandberg-Larsen et al., 2008).

As all my results are based on sibling comparisons and mothers who change their behaviors *between* pregnancies identify these effects, I examine potential challenges to my identification strategy and carefully examine changing and non-changing mothers to gain insights in the driving forces behind mothers' behaviors. Crucially, mothers' employment status should not be driven by either learning effects or unobserved time-varying factors that also impact their children's outcomes. While most sibling studies leave this "black box" of what drives behavioral changes (e.g. in prenatal smoking behavior) unaddressed, several informal tests can help to shed light on the reasons for mothers' decisions. I perform the following ones: I examine changing and non-changing mothers on a set of observable characteristics, I examine the correlation of mothers' behaviors and outcomes in their first observed pregnancy and behaviors in later pregnancies, and I look at subsamples of mothers—e.g. mothers with closely spaced children. My conclusions from these tests are the following: first, mothers who change their employment situation between pregnancies also change on other dimension, e.g. many "starters" increase their education. Thus my rich set of control variables is a major advantage. Second, mothers' first pregnancy outcome is not correlated with their second pregnancy employment decision (conditional on other observables). While this finding has no causal interpretation, it indicates that learning effects do not appear to be very important. I take my finding of similar employment effects for mothers with closely spaced children and of similar employment effects based on only starting or quitting mothers (with respect to employment) as another argument for this claim. I thus argue that conditional on a large set of controls and the mother fixed effect, I identify the effect of maternal work on birth outcomes.

The rest of the paper unfolds as follows: Section 2 discusses relevant literature and presents necessary background. Section 3 outlines the empirical methods, and section 4 describes the data. Section 5 presents my results and discusses their robustness. Section 6 concludes.

2 Literature and Background

While the economics literature on the effect of parental inputs for child health at birth has focused on a number of factors, including maternal age at birth (Rosenzweig and Wolpin, 1995), maternal education (McCrary and Royer, 2011; Chou et al., 2007; Currie and Moretti, 2003; Lindeboom et al., 2009), family income (Case et al., 2002), prenatal maternal smoking (Abrevaya, 2006; Lien and Evans, 2005; Abrevaya and Dahl, 2008; Currie et al., 2009), and the use of prenatal care (Rosenzweig and Schultz, 1983; Dave et al., 2008), maternal employment during pregnancy has not been researched extensively.

In economics, maternal employment has mainly been studied in a post-natal context focusing on the effect of maternal leave taking on early life health (Ruhm, 2000; Rossin, 2011) and longer-run outcomes such as educational attainment (Rasmussen, 2010; Dustmann and Schönberg, 2008). Ruhm (2000) uses cross-country variation in maternity leave programs and finds positive association of maternity leave after birth with aggregated child health measures. While Rossin also includes birth outcomes into her analysis of maternity leave enactment, she points out that postnatal leave drives her main results for the U.S. She finds a positive effect of the introduction of unpaid maternity leave on birth outcomes and infant mortality for mothers who can take advantage of the maternity leave, i.e. who can afford to leave the labor market. Turning to longer-run benefits of maternity leave, Dustmann and Schönberg (2008) and Rasmussen (2010) find no significant long-run effect of mothers' leave taking.

A second set of studies has studied the effects of maternal return to employment in the first years of the child's life. This literature has yielded mixed results with some studies finding negative effects of employment in the first year after childbirth on child cognitive outcomes (e.g., Brooks–Gunn et al., 2002; Ruhm, 2004) and other studies finding little evidence for negative effects (Baker and Milligan, 2010). Highlighting the importance of context, a recent study from Denmark finds no effects of maternal employment later in the child's life for measures of child obesity (Greve, 2011).

The above papers refer to a number of potential mechanisms for an effect of postnatal maternal employment on child outcomes. Negative effects potentially are due to time constraints and stress on the part of the mother and the poor quality of childcare arrangements. As for offsetting factors, researchers have argued that income effects could play an important role explaining positive effects for maternal employment, as well as high-quality childcare arrangements (Datta Gupta and Simonsen, 2010).

Most studies on maternal *prenatal* employment come from the field of occupational health, which has traditionally primarily focused on risk factors such as the exposure to a hazardous workplace (for an overview Gabbe and Turner, 1997). In more recent years, this literature has shifted its focus to also examining other potential channels for the effect of maternal employment, such as stress and anxiety. General stress during pregnancy has been related to lower birth weight and premature birth in some studies while others have not found any correlation (Copper et al., 1996; Guendelman et al., 2009; Hedegaard et al., 1996). As existing studies in occupational health mainly use cross-sectional data, the causal effect of employment during pregnancy is not well understood. Due to their design that ignores endogeneity concerns, many studies are prone to find a "healthy worker effect" (Gabbe and Turner, 1997), i.e., a positive employment effect due to selection of healthy mothers into employment.

In their paper on intrafamiliar resource allocation, Del Bono and co-authors are the first economics study to incorporate information on maternal prenatal employment into the analysis (Del Bono et al., 2008). Exploiting within-family variation, they find positive effects for work interruptions during pregnancy. They attribute this finding—and the finding that low-educated mothers benefit most—to the impact of negative stress on the part of the (working) mother. They also highlight the importance of institutional differences between the two countries, such as the design of maternity leave.

Nevertheless, other recent empirical work raises doubts about the general validity of the stress channel. While Aizer et al. (2009) find important negative effects for the exposure to stress in utero for child health at birth, cognition and health at age 7, they find no correlation of stress levels and prenatal employment in their sample. Identification in this paper relies on the comparison of siblings exposed to different levels of cortisol, i.e. different levels of maternal stress. As this paper also points out, it is not straightforward that maternal employment during pregnancy should generate stress per se.

Stress could also be a result of exclusion from the labor market rather than employment. A number of studies—many of them investigating data from Scandinavia—have investigated whether there are negative effects of unemployment on various health outcomes. Facing similar identification problems relating to health selection and reversed causality, some studies have found a negative effect of displacement from work and unemployment on health outcomes while others show no effects (Browning et al., 2006; Kuhn et al., 2009). Studies that have examined the societal costs of plant closures—and thus rely on credibly exogenous variation of employment status—have found important negative health effects for displaced workers and have attributed this finding to stress imposed by job loss (Sullivan and von Wachter, 2009). Consequently, if exclusion from the labor market due to e.g. unemployment imposes stress on the pregnant woman, we could observe a positive employment effect for birth outcomes. Furthermore, employment connects women to workplaces and other women at these workplaces, which in turn might influence both their decisions during pregnancy and even the timing of their fertility decisions (Hensvik and Nilsson, 2010). Thus peer effects might be a channel for positive effects of employment on birth outcomes.

As a consequence, the effect of maternal employment during pregnancy on birth outcomes is an empirical question that requires careful consideration of heterogeneity between mothers and the endogeneity of the employment decision of pregnant women. Denmark makes an interesting case for studying employment during pregnancy. Prenatal care is universal and free of charge in Denmark. Thus while the income channel has been highlighted in studies on postnatal employment from the U.S., it should be less relevant in this context. For Denmark prenatal employment should not improve mothers' access to medical care during pregnancy by increasing income.

While most Danish women work during pregnancy, a number of factors make Denmark different to, e.g., the U.S.: Employers are legally obliged to accommodate work tasks to the needs of pregnant women. This feature of employment legislation aims at keeping pregnant women on the labor market by accommodating their needs and without endangering the health of the pregnant woman and the unborn child. Thus especially in manual jobs, tasks are most likely accommodated to the pregnant woman's needs.

Furthermore, opposed to the U.S., Denmark has one of the most generous parental leave systems in the world. In the period considered in this paper, all women are entitled to four weeks of prenatal maternity leave and close to all women are entitled to a maternity leave benefit on the level of the unemployment benefit with a benefit cap. Furthermore, many women are covered by additional collective agreements that grant them their full wage during the prenatal maternity leave.[2]

A final and important feature of the Danish labor market regulation is sick leave. Employed pregnant women with health problems can get sick listed by their general practitioners (GP). GPs can sick list women due to pregnancy complications that could endanger her health or the health of the baby. Women who are sick listed (and meet an employment criteria) are eligible for a benefit during leave, either temporary or for the rest of their pregnancy. Recent figures show that pregnant women in Denmark on average are absent from the labor market 48 days (including the 4 weeks of pre-birth leave) (Danish Ministry of Employment,

[2]Not only previously employed women but also women on welfare benefits are entitled to paid prenatal maternity leave. The very few exceptions from paid prenatal maternity leave include students (who in turn receive a higher student benefit after childbirth) and women who are not entitled to any welfare benefits, i.e., home-makers or uninsured self-employed women. Those two groups are very small. Post-birth parental leave schemes are generous as well and grant all parents access to 14 weeks of maternity leave and up to 32 weeks of parental leave. The duration of paid postnatal leave is also subject to collective bargaining for an overview see (for an overview see Datta Gupta et al., 2006)

2010). This number covers over differences according to mothers' characteristics (ibid). In my analysis I control for mothers' sick listing during pregnancy to ensure that my estimate for maternal working during pregnancy does not measure the mother's formal employment status irrespectively of whether she actually works. Thus my estimate for maternal employment distinguishes the potential effects of holding a job from the effect of actually being at work.

3 Empirical Methods

To identify the causal effect of prenatal inputs for child health, economists have turned to one of two strategies. Instrumental variables have primarily been used for estimating the effect of prenatal maternal schooling or smoking on birth outcomes. However, finding relevant and valid instruments for maternal inputs has proven to be difficult and assessing the impact of several inputs requires more than one instrument at a time. Therefore, a small number of studies have exploited either variation among mothers who are twins or siblings (Currie and Moretti, 2007; Rosenzweig and Wolpin, 1995) or variation in consecutive births to the same mother (Abrevaya, 2006; Del Bono et al., 2008; Rosenzweig and Wolpin, 1995; Currie et al., 2009).

The model that describes the association of birth outcomes and maternal inputs—among them maternal employment—takes the form

$$Y_{i,s} = \alpha \times EMP_{i,s} + \beta \times X_{i,s} + c_s + \epsilon_{i,s} \tag{1}$$

for child i in family s where $Y_{i,s}$ is birth weight or the likelihood of premature birth. $EMP_{i,s}$ is an indicator for maternal employment during pregnancy with child i measured around 8.5 weeks before the mothers' due date or in pregnancy week 31.[3] $X_{i,s}$ is a vector of observable health inputs and control variables. $\epsilon_{i,s}$ is an idiosyncratic error term. Estimating this model by OLS on pooled data for all mothers and their children leads most likely to biased results in the presence of c_s, which stands for (time-invariant) unobserved variables common to all children of the family and correlated with both inputs and the dependent variable. Prime candidates for omitted variables are maternal preferences, abilities or underlying health.

[3] Accounting for the exact week for the second interview does not change my results. Results are available on request.

To solve the problems induced by mother-specific and time-invariant unobservables, I turn to sibling data and estimate the child health equation in a differenced form for consecutive children i and j to the same mother as in (2).

$$(Y_{i,s} - Y_{j,s}) = \alpha \times (EMP_{i,s} - EMP_{j,s}) + \beta \times (X_{i,s} - X_{j,s}) + (\epsilon_{i,s} - \epsilon_{j,s}) \tag{2}$$

As identification in this paper is based on variation between siblings born to the same mother, at least three concerns remain: unobserved time-varying and correlated behaviors, measurement error, and dynamic parental response.

While the mother FE takes care of bias induced by unobserved time-invariant mother characteristics, the estimation will only yield unbiased estimates if no correlated and time-varying variables are omitted from the estimation. Earlier sibling studies that have focused on prenatal maternal smoking usually find smaller estimates when taking a mother FE into account (see references in Abrevaya, 2006). However, this attenuation could also be driven by both measurement error or omitted maternal behaviors. In a multivariate framework it is hard to access the direction of bias induced by omitted variables, which is dependent on the size of the effect of the omitted variables and their covariance with all included inputs.

Mothers who change their health behaviors between births could do so with respect to other unobserved behaviors – either "turning healthy", i.e. choosing bundles of healthy or unhealthy behaviors, or "substituting one evil with another", i.e. engaging in one harmful behaviors while stopping with another. Abrevaya (2006) finds that if mothers change smoking behavior, they also change other behaviors "in the same direction". However, his set of available health behavior variables is very restricted. Given that I have a rich set of controls for maternal inputs at my disposal, I am confident that time-varying confounding factors are less important. Moreover, as this paper includes an array of controls for maternal behaviors in its analysis, it can examine whether earlier estimates based on sibling comparisons—especially for maternal smoking—are robust to this inclusion of omitted time-varying factors. Section 4 examines the pattern of changes in maternal health behaviors in greater detail.

Another concern in sibling studies is the impact of measurement error. Attenuation bias induced by classical measurement error is typically bigger in FE models than in OLS (Griliches, 1979). However, while often assumed to be, measurement error is not necessarily classical. Recall and justification bias can be induced when mothers report health behaviors and misreport systematically. By the late 1990s, for example, first-time mothers had been exposed to a considerable amount of information about the negative effects of smoking. Thus if mothers misreport smoking, it is very likely that they understate their smoking behavior. Similarly, the Danish norm of working as a mother could induce women to over-report employment during pregnancy. The FE approach helps minimising the impact of this systematic misreporting if we assume that mothers misreport similarly in consecutive pregnancies. Furthermore, the data generation process with timely interviews and the additional controls from the administrative registries make me confident that I face less measurement error in my analysis than earlier studies.

A final concern is the potential bias induced by violation of the exogeneity assumption, i.e., parental dynamic investments in consecutive siblings. For the FE estimation to identify the effect of maternal employment, employment in the second observed pregnancy should not depend dynamically on the outcomes of the first pregnancy. As this assumption can be questioned in the analysis of prenatal inputs in child health, Del Bono et al. (2008) have focused their analyses on identifying mothers' dynamic investments. They propose a model where mothers react on children's idiosyncratic endowments and adapt their inputs in following pregnancies accordingly. However, applying a FE instrumental variable approach, they find surprisingly stable results for their estimates for maternal inputs. Thus their main results are not sensitive to relaxing assumptions about the nature of parental response. While this paper does not resolve the question of dynamic response, I examine mothers' behaviors in consecutive pregnancies carefully to examine the potential impact of learning between pregnancies. Additionally, I use an instrumental variable approach with cross-sectional data to test the robustness of my conclusions based on sibling comparisons to an alternative set of identifying assumptions. I present the results in section 5.1.

4 Data and Summary Statistics

This paper uses data from two sources: first, the "Danish National Birth Cohort" (DNBC) is a survey, which comprises information on pregnancies and births from 1998 through 2003.[4] Second, I link this data to administrative register data on the universe of singleton live births in Denmark for the same period.

[4] I drop survey year 1997 as I only have information on 29 children from this year in my data.

The use of the DNBC survey has two advantages: First, the survey data has been collected explicitly for studying the determinants of health at birth and early child health. Thus the data contains information on an array of health behaviors likely to have an impact on children's health at birth which allows me to control very accurately for observable differences between mothers. While focusing on maternal employment, this study is the first to examine the impact of maternal employment, smoking and alcohol consumption simultaneously in a sibling framework.

Second, the DNBC survey was administered to mothers during their pregnancy. Therefore, the data does not suffer from recall and most likely less justification bias, both of which can constitute a considerable problem in retrospective reports on inputs after birth (Currie, 2000). Mothers were invited to answer two surveys, one around week 12 and one around week 30 of their pregnancy. Furthermore, the linked administrative register data adds objective outcome measures to the analysis. Given the national personal identifier and the reliability of the administrative register data, this paper does not encounter problems of potential mismatches of parents and children and I expect a very low level of measurement error in left-hand side variables.

The DNBC survey covers about one third of all detected pregnancies in Denmark in the period considered here.[5] It contains information on a specific birth outcome for 100,309 pregnancies. Merging the survey data with the administrative registers and considering non-response and attrition, my final sample consists of 76,281 children for whom mothers have at least answered pregnancy interview two at week 30 of their pregnancy. Of those, 10,194 children are part of the sibling sample. The sibling sample was not constructed on purpose but is the result of mothers participating with more than one child during the rather large sampling window.[6]

Table 1 presents summary statistics for the sibling sample, the full survey sample, and all singleton births in Denmark for 1998-2003.[7] All outcome measures of this paper come from

[5]For the data collection, GPs invited pregnant women at their first doctor visit to participate in the study, i.e., the sampling unit is pregnancies. As a result of the sampling scheme, half of the non-participation in the first pregnancy interview results from a lack of GP cooperation, while the other half is attributable to women's non-response (Nohr et al., 2006).

[6]Of the children in the sibling sample, under 200 children in sibling pairs do not share the same father. This time-varying factor could be important in interpreting the mother FE as capturing family-specific endowment. However, controlling for this factor and estimating the model excluding those children does not change the results of the analysis.

[7]The survey samples contain children with at least information from pregnancy interview two, which is

the Danish administrative registers. Birth weight is measured in gram and fetal growth is defined as birth weight divided by gestational age (i.e., it accounts for the fact that many low birth weight babies are born prematurely.) While comparable with respect to birth weight and gestational length, the survey samples have a lower percentage of premature birth than the overall population with around 3.5 percent for all survey mothers and 3 percent for the sibling sample, respectively. Premature birth is defined as birth before completed 37 weeks of gestation.

Using register data to compare mothers in the survey to all mothers with singleton births in the period, the table shows that survey mothers are somewhat positively selected from the general population. They have higher employment rates in the year proceeding their birth. 88 percent of survey mothers are employed as opposed to around 80 percent in the source population of mothers. Furthermore, mothers in white-collar occupational groups (ISCO groups one to four) are overrepresented in the survey. Survey mothers have lower smoking prevalence rates during pregnancy, they are more likely to cohabit and have higher educational attainment than the overall population.

For child's sex and parity, and mother's age at birth, the survey compares well to the overall population. With respect to pregnancy-related health, the survey women also resemble the overall population. The indicator for pregnancy complications includes a set of health conditions; among them diabetes, preeclampsia, edema, and placenta praevia.[8]

As my analysis exploits information on mothers with more than one child, Table 1 additionally compares mothers who participate in the survey with more than one child and all survey mothers. Here I focus on reports from the DNBC survey. An identical percentage of mothers in both samples report that they work at both pregnancy interviews. Moreover, working mothers in the sibling sample report similar occurrence of sick listing from their job.

The employment figures for Denmark are higher than the ones reported in Del Bono et al. (2008) for the UK and the U.S. As a point of reference, in their UK and the U.S. samples between 32 and 50 percent of women, respectively, are not employed at all during their pregnancy.[9] The percentage of women who report that they work is 81 and 77 at the first and second pregnancy interview of the DNBC. Taking into account that some of these

my main data source. While a small number of mothers has not completed interview one, attrition between interview one and two is more pronounced (see section 5.1 for further details).

[8]The diagnoses included are the ICD10 codes DO11-16, DO24, DO30-48.

[9]However, as mentioned earlier, their U.S. figures are older and thus not comparable to more recent years.

women only formally hold a job but are sick listed at the time of the second interview, around 66 percent of DNBC mothers work at interview two.

The average working time for working mothers in the DNBC at interview one is around 35.5 hours (unfortunately, this measure is only available at the first pregnancy interview.) This figure is somewhat lower than full time but slightly higher than the mean reported in the National Labor Force Survey (Danish Ministry of Employment, 2011).[10] Women who worked around interview one but have stopped working before interview two have on average only worked 3 weekly hours to begin with. Thus employment during pregnancy seems to be rather stable for Danish women, i.e. women either work *throughout* their pregnancy or they do not work at all. A way of validating the mothers' reports on employment is to compare the survey to administrative register data. As Table 1 shows, the employment figures from the survey and the administrative registers appear credible—the percentage of employed mothers in the year prior to birth is as expected higher than in the survey.[11]

For a couple of other characteristics, the full survey sample and the sibling sample reveal differences: At interview one, around 20 and 26 percent of the mothers in the sibling and full survey sample report they have smoked during their pregnancy or are still smoking.[12] At interview two, these figures have decreased to 11 and 16 percent, respectively. To validate survey reports for smoking, I also compare survey and register data: the percentage of women reporting smoking to their midwives at birth is literally identical to the percentage of mothers who report smoking at the second DNBC interview.

For alcohol consumption, the percentage of mothers who report that they drink at least one unit of alcohol per week increases between the two pregnancy interviews from 26 and 24 percent to 33 and 31 percent for the full sample and sibling sample, respectively.[13] Mothers who drink consume on average around two units of alcohol per week. The increase in alcohol

[10]Between 2000 and 2003 the average weekly working hours for employed women in the National Labor Force Survey was around 33 hours (Danish Ministry of Employment, 2011).

[11]The employment status is registered in November each year.

[12]Danish women converge in behavior with women in countries such as the UK or the U.S. In the 1980s 38 and 41 percent of pregnant women in two major Danish towns smoked during pregnancy (Olsen et al., 1989). The figures in the first DNBC interview is close to UK smoking figures and considerably higher than those reported for the U.S.(Del Bono et al., 2008). At the second interview, the Danish figure is closer to those reported from the U.S. (Abrevaya, 2006). The DNBC data reveals similar associations of maternal smoking with observable characteristics such as age or educational group as found in earlier studies. The youngest and oldest mothers have the highest percentage of smokers and women at the mean age for first-time mothers having the lowest. For smoking throughout pregnancy, the percentage of women in the lowest educational groups is highest.

[13]One unit equals 1 bottle of beer, 1 glass of wine, or 1 glass of liquor.

consumption between interviews likely reflects women's expectations about critical periods with respect to alcohol during pregnancy. Mothers might suspect alcohol consumption to be more damaging in the early stages of pregnancy.[14] However, among researchers no consensus exists about the importance of timing of alcohol consumption during pregnancy (see e.g., Henderson et al., 2007), although the kind of damage to the fetus might vary between trimesters (for an overview see Nilsson, 2008). Furthermore, variation in national guidelines also reflects uncertainty about the impact of low but regular doses of alcohol consumed during pregnancy (O'Brien, 2007; Nathanson et al., 2007).

Taken together, the data from the DNBC survey illustrates that mothers change a number of health behaviors considerably *during* pregnancy. This finding underlines the importance of the timing of measurement. Although the comparison of health behaviors at different stages of pregnancy has revealed considerable changes, in the main analysis I focus on information from pregnancy interview two. However, in my robustness checks I also test specifications with information from interview one. These tests shed light on dose-response relationships for health behaviors vs. critical periods for certain behaviors such as alcohol consumption. Furthermore, the tests examine whether women selectively leave the survey (i.e. whether working women are more or less likely to answer the second pregnancy interview).

Changes *during* pregnancy are interesting in its own right. However, given that they identify the effect of maternal employment in the sibling framework, changes *between* pregnancies are crucial for the validity of the strategy of this paper. Thus I discuss several challenges to identification in this paper and perform informal tests to examine their importance.

One first challenge to identification could arise from omitted time-varying and correlated changes in behaviors between pregnancies. The availability of information on a range of health behaviors allows me to examine the question as to whether mothers change other health behaviors as they change empoyment status *between* pregnancies. This question has been discussed in the context of sibling studies that only have few controls for time-varying behaviors at their disposal. Table 2 divides mothers according to their employment status at the second pregnancy interview in the first and second observed pregnancy: starters only work in their second pregnancy, quitters only work in the first pregnancy. Always workers

[14]This possibility is supported by the finding that the women in the sample on average drank even more before their pregnancies. The percentage of mothers who report consuming alcohol increases with age and for higher educational attainment.

and never workers do not change their behavior between pregancies. The table compares mothers' birth outcomes and other observable characteristics that change over time.

All groups have higher birth weight means for the second observed pregnancies. However, as the percentage of premature babies per group illustrates, not all birth outcomes improve for higher parity children in all groups. The percentage of mothers with a pregnancy complication is higher for all groups in the first pregnancy. These findings suggest that changes in employment between births are not driven by differential health trends for mothers in the different groups and their first (observed) child. Table 2 also shows that mothers in the starter group are the youngest in their first pregnancy and increase their education significantly between births. This observation suggests that many of these women have their first child while still in the educational system. This behavior is common in Denmark. While consuming on different levels, the percentage of women smoking and consuming alcohol in the groups with different employment patterns changes surprisingly similarly: The percentage of smoking mothers decreases and the percentage of mothers who consume alcohol increases between pregnancies in all groups. This finding suggests that controlling for these behaviors should not matter much for the estimate of the employment effect.

Another concern in the sibling framework arises from dynamic parental behviors: If mothers change employment status between births due to their learning from the outcomes of their first birth, the sibling comparison is not valid. Appendix Table A.1 presents an informal test of mothers' changing behaviors and its determinants. The table shows regression results for second pregnancy employment on birth outcomes and behaviors in the first pregnancy. I should not find an association of first pregnancy outcomes and second pregnancy employment as this finding would strongly suggest that mothers change their behavior due to learning. Although these estimates have no causal interpretation, the results in Table A.1 suggest that conditional on mother observables the outcome of the first pregnancy—as measured in this paper—is not a good predictor for second pregnancy employment status. Important predictors for second pregnancy emloyment—which I also include as controls—are white-collar occupational status and maternal sick leave in the first pregnancy. Furthermore, the coefficient for first pregnancy employment as expected suggests that women who work in the first observed pregnancy are very likely to also work in pregnancy two.

A final concern for my strategy of comparing the outcomes of siblings is attrition of mothers before the interview around week 30 of the pregnancy. If, e.g. working mothers experience more premature births that happen before interview two, these mothers will by construction be missing in the second interview and drive a positive employment effect (as remaining employed mothers have a lower probability of preterm birth). However, as I show in section 5.1 on robustness tests, the event of extreme premature birth (birth before week 30) is very rare. Furthermore, by using birth outcome data from the administrative registry for all children and data from the survey about the timing of the (attempted) second pregnancy interview, section 5.1 shows that working mothers have no higher probability of very preterm birth and that working at the first interview is no good predictor for missingness at the second interview. It concludes that the main results cannot solely be attributed to employed/unemployed mothers leaving the survey before the second interview.

5 Main Results

Table 3 presents my baseline OLS estimates for birth weight and the probability of premature birth. As results for fetal growth confirm the birth weight results, I present them in Appendix Table A.2 for convenience. I estimate the OLS models on both the sibling sample—columns one through four—and full survey sample—columns five through eight. Table 3 presents two models for both premature birth and birth weight, one narrow control and one extended control model.

The narrow control model includes information on maternal educational level (in four levels and an indicator for missing educational level), maternal age (an indicator for maternal age over 25)[15], cohabitation status at birth, child sex and whether the child is the first-born, and a set of year dummies. Furthermore, I include only two maternal health behaviors: an indicator that turns one if the mother is formally employed (not accounting for being sick listed from work) at interview two and an indicator for smoking at interview two.

In the extended control model, I add further control variables for maternal behaviors, thereby giving a first impression of the sensitivity of my estimates to potential omitted variables. I account for several observable differences between women that have been shown to correlate with birth outcomes: I include an indicator for maternal sick leave during pregnancy (more than three days) and pregnancy-related health problems – these indicators capture

[15]I have tested various specifications for the age variable and find my main coefficients of interest unchanged.

pregnancy-related health problems that potentially lead women to stop working and introduce reversed causality to my analysis. Furthermore, controlling for sick leave means that the work indicator measures actually being present at the workplace. The model also controls for indicators for the mother being in a blue-collar and white-collar occupational group (missing occupation is the reference group), whether the mother smoked, whether she exercised and consumed fish. Additionally, I account for the number of daily cigarettes, the units of alcohol per week and the pre-pregnancy BMI. All columns report standard errors that are clustered at the mother level.

The OLS results indicate that maternal employment in the last trimester decreases the risk of premature birth in both samples and the coefficient is bigger when other maternal health behaviors are taken into account. However, mothers who are sick listed from work experience no positive employment effect—as expected, as they leave work due to health problems that most likely affect their children's health, too. Maternal employment has no effect for birth weight. The OLS suggests a decrease of the probability of premature birth of around 1.4 percentage points for working mother in the sibling sample.

As expected, I find that pregancy-related diagnoses are related to both an increased risk of premature birth and reduced birth weight. As the indicator for pregnancy complications covers over a range of diagnoses, the modest size of the coefficient should be interpreted with caution. For other controls two main findings emerge: first, at the mean of around 8 daily cigarettes, I find that mean birth weight decreases by 228g for the sibling sample. This figure compares well to OLS results found elsewhere and to the OLS estimates without detailed control (Abrevaya, 2006; Currie et al., 2009). Second, at the mean of 2 weekly units of alcohol consumption displays a very small negative effect on birth weight with a reduction of around 24g—the estimate in the full survey sample is even smaller at around 10g. This finding seems to confirm the popular notion of small amounts of alcohol not being hurtful to the health outcomes considered here.[16] Furthermore, the differences in estimates for the sibling and full sample highlight another crucial point present in all sibling studies: estimates are based on a select sample of mothers with more than one child and results might not be comparable to estimates for the full population of mothers.

At the same time, the OLS estimates have to be interpreted cautiously as heterogeneity

[16]While maternal exercise and nutrition have been subject to considerable debate in the scientific and public sphere, their effect on the birth outcomes considered here are of minor importance.

between mothers—both in the sibling and full sample—could bias the estimates. Thus Tables 4 and 5 present the results of an analysis based on comparisons between siblings born to the same mother. The tables present results of this mother FE model for premature birth, birth weight and fetal growth and include the narrow and extended set of controls, respectively. Table 5 does not control for the number of daily cigarettes, which does not contribute significantly to the estimates of the smoking effect. Standard errors are clustered at the level of the mother in both tables.

A first observation is that the FE estimates for maternal employment are bigger in absolute size than the OLS estimates, i.e. the protective effect of employment—e.g. through less stress and peer effects at the workplace—is underestimated in the cross-sectional model. A common finding in the literature on, e.g., the impact of maternal smoking, is that OLS models overestimate the effect of this behavior. Just as in the case of maternal employment it appears intuitive to assume a positive correlation for "high quality mothers" unobservables with both positive birth outcomes and positive health behaviors.

Several reasons for the present and somewhat counter-intuitive finding for maternal employment are possible. While the OLS exploits variation between all mothers, the FE model is identified from mothers who change employment status between births. The latter mothers are potentially the ones that profit most from the stated mechanisms of reduced stress and networks at work. To analyze this point further, I look at mothers from different educational and occupational groups separately in the following robustness tests.

Furthermore, another reason for the bigger FE results could be that they correct for the impact of systematic measurement error. As an example, Smith (2009) finds bigger sibling-based estimates for the effect of early life health on later life socio-economic outcomes. He argues that systematic misreporting between individuals is smaller for siblings (who share family environment). In my example this would translate to mothers misreporting their employment status—unemployed mothers would be prone to state that they are employed—and the FE correcting for the bias induced to the between-mother estimates by this form of measurement error.

Finally, although the claim that mothers' unobserved time-invariant characteristics capture overall quality of the mother that is positively correlated with both employment and child outcomes appears plausible, this assumption can be challenged. Currie et al. (2009) illustrate by the example of pollution and residential choices that “high quality” mothers could be either more likely to move to low-polution areas (e.g., because they have higher income and preferences for less pollution) or to be exposed to more pollution (e.g., because they value inner-city residence). Similar mechanisms related to working women’s preferences could be at work here.

The main FE results in Table 5 confirm a protective effect for maternal employment with respect to premature birth. Employed mothers are 2.8 percentage points less likely to deliver premature. This employment effect is large and translates into a 93 percent decrease of the probability of premature birth for working mothers. As a comparison between Tables 4 and 5 shows, the included control variables do not matter for the size of the employment effect in the FE model. There is no birth weight effect of maternal employment in the FE model, the coefficient is very small and insignificantly estimated. In contrast, being diagnosed with a pregnancy-related health condition decreases birth weight and fetal growth, and it increases the risk of premature birth. Its effect on the probability of premature birth is—similar to the effect of maternal work—very large (a 2.5 percentage points increase), making the two factors the ones that matter most for this outcome.

I also find smoking results of conventional size for birth weight and fetal growth and in line with earlier sibling studies, the size of the coefficient is smaller than in the OLS results: Smoking decreases birth weight with around 95g and fetal growth by around 2.4g/week. Consequently, this finding shows that earlier sibling studies on the effect of maternal smoking are not biased by omitted variable though lacking many of the control variables available in this study.

A rarely available measure of maternal behaviors during pregnancy is maternal alcohol consumption. Although it does not change the size of the smoking or employment effect, the inclusion of this maternal health behavior is interesting in its own right. The effect of alcohol consumption is bigger in the FE models when compared to OLS: At the mean consumption (2 units/week), mothers are around 1.6 percentage points more likely to experience a premature birth and birth weight is reduced by around 36g. This effect indicates that also moderate alcohol consumption during pregnancy should be discouraged by health guidelines. In 99 percent of all pregnancies in the sample the mother has consumed less than four units of alcohol per week and in 70 percent of all pregnancies less than two. Excluding the mothers who have had more that ten weekly units, the alcohol result remains unchanged—an indication that it is not driven by high-consumption mothers.[17]

Turning to the question of heterogeneous effects of prenatal employment, Table 6 presents FE estimates for subgroups of mothers according to their educational group and occupational status at the first pregnancy observed in the data. For convenience the table only shows the results for premature birth and birth weight. Appendix Table A.3 shows the results for fetal growth. Dividing my sample of mothers according to educational level (only high school vs higher level of education than high school) and occupational status in the year prior to birth gives insights into potential heterogeneous effects according to mothers' characteristics.[18] As columns one through four in Table 6 show, mothers with high and low educational attainment (in pregnancy one) display similar employment effects for premature birth and birth weight. Birth weight results are imprecisely estimated and confirm the absence of important birth weight effects.[19] This finding is informative in the Danish context: many of the mothers who change employment status between births start out as students in their first pregnancy and enter the labor market before they have their second child. As employment effects are not different according to mothers' educational group during their first pregnancy, I take this result as another indication that the income channel—which would be most important for students entering the labor market—is not the driving force of my results.

Columns five through eight divide mothers by their occupational status the year before their first observed pregnancy. While I find that maternal employment reduces the risk of premature birth for all mothers, the effects are bigger for blue-collar occupation mothers in employment. While birth weight results in Table 6 are only significantly estimated for white-collar mothers, they are small and resemble the main results. As my analysis controls for maternal sick listing from the job, the bigger results for blue-collar workers could highlight the importance of workplace accommodation, which is more likely to occur for blue-collar workers. Also blue-collar mothers could benefit more from actually being at their workplace, e.g. through peer effects.

Table 6 also reveals an interesting pattern with respect to smoking and alcohol consump-

[17]Estimating the models with indicators for more than one, more than two, and more than three weekly units respectively, confirms the negative alcohol effect.

[18]The table excludes mothers with missing occupation or education, respectively.

[19]Both groups have enough changers out and into employment between pregnancies; 7 and 3 percent for the low educational group and 4 and 5 percent for the high educational group.

tion and their effects on the considered birth outcomes: While the negative smoking effect for birth weight is driven by mothers with low education, the alcohol effect is strongly significant and bigger for highly educated mothers. Smoking and alcohol effects vary in a comparable way by occupational group, with white-collar occupation mothers displaying a strong alcohol effect and blue-collar occupation mothers displaying a negative smoking effect. At the relevant subsample mean for high-education and white-collar mothers of around 2 units of alcohol per week, alcohol consumption increases the risk of premature birth significantly by 2.2 and 1.8 percentage points, respectively.

This finding could reflect differences in consumption patterns and could suggest that while high educated mothers are—as earlier found—less likely to engage in smoking during pregnancy, they are less concerned with the potential harmful effect of (the occasional glass of) alcohol. Furthermore, heterogeneous smoking effects for mothers with different educational status have earlier been explained with higher education mothers being able to better offset negative smoking effects by more productively choosing other health inputs and thus compensating for the negative effect of smoking. However, the negative alcohol effect for high educated mothers casts doubt on this hypothesis.

5.1 Robustness Tests

In this section I present a number of informal tests for the robustness of my results. While ultimately unobserved to me, these tests can also shed some light on the question as to why mothers change employment status between pregnancies. I start by examining the potential threat that attrition of working mothers who have a very premature birth before interview two drives my results. I continue by examining challenges to the assumptions made in the FE model.

My main analysis focuses on mothers who complete both pregnancy interviews of the DNBC. Especially for the outcome premature birth, it is crucial to rule out that selective attrition drives the positive employment results. The timing of interview two—scheduled for around week 30 of the pregnancy—means that mothers who experience a very premature birth—before week 30—are excluded from the second survey round by construction. Table 7 shows the outcome of very premature birth for all Danish mothers and survey mothers. It uses administrative register data on birth outcomes that is available for all mothers irrespectively of survey participation. Crucially for my strategy, the percentage of very premature birth is generally very low in Denmark and my survey samples. Thus these observations should not drive my results for premature birth.[20]

While the second DNBC interview was *scheduled* around week 30, many women were interviewed later due to time constraints and follow-up problems. If these differences in the timing of the second interview were related to women's employment, they could bias my estimates, i.e., if working mothers had a premature birth before they participated in the survey, the positive effect of employment could be due to this differential trend.

Table 8 shows regressions for the probability of being pregnant at interview two (i.e., answering at least the introduction question of the second DNBC interview) and the probability of dropping out of the DNBC completely before interview two. The table shows that an indicator for whether the second interview was conducted before the expected due date (calculated at interview one) is the best predictor for whether the mother is still pregnant at interview two. Employment and other mother characteristics do not predict whether the mother is still pregnant and thus participates in interview two (conditional on interview one). As not all interviews could be performed in a timely fashion, some of the non-response in the survey is due to the timing of interview. This variation in timing of interview is unrelated to observed mothers' characteristics at interview one.[21] Furthermore, Table 8 shows that employment at interview one—when accounting for mothers' characteristics at interview one and a mother FE—does not predict complete missingness of information for interview two (i.e., no follow up at all). Thus it suggests that working or non-working mothers do not leave the sample selectively.

Further testing the impact of mothers leaving the survey before the second interview, Table 9 shows estimates for the effect of maternal behaviors at interview one on the probability of very premature and premature birth (here I can include mothers who only answer interview one). These regressions also allow me to examine the importance of timing of measurement of health behaviors during pregnancy. Columns one to four report results for all mothers with an completed first pregnancy interview. Employment at interview one does not predict very premature birth and the results for premature birth resemble my main results (for both OLS and FE models). This finding is not surprising as changes in employment status during

[20] I have no information on fetal death/spontaneous abortion very early during pregnancy.
[21] Results for unrelatedness of timing and observables at interview one are available on request.

pregnancy are not as common in the Danish context as illustrated in the summary statistics. In contrast, smoking and alcohol results based on measures from interview one are smaller and mainly insignificantly estimated. The changes in these two maternal behaviors during pregnancy are big: a considerable percentage of mothers stop smoking and start consuming alcohol during pregnancy (as illustrated in Table 1). Consequently, my results in Table 9 suggest a dose-response relationship of these two maternal behaviors and the birth outcomes considered.

Finally, in columns five to seven Table 9 presents an analysis using data from interview one and imputed employment status for all mothers with missing information from interview two (either no follow up or no pregnancy at interview two). Column five presents results using data for women with employment information from interview two, in column six all imputed values for missing employment are set to one and in column seven all imputed values are set to zero. The table confirms the main results for maternal employment: Employment reduces the probability of premature birth, results for the imputed employment measures are similar to the main results. Taken together, my findings suggest that the main results for the protective effect of maternal employment on premature birth are not driven by selective attrition of mothers between survey waves.

As my main results take mother-specific and time-invariant factors into account, I can rule out that factors such as ability or taste for employment bias my results. Furthermore, my control for a broad set of health behaviors and maternal characteristics make it more credible to argue that time-changing omitted variables do not bias the estimates. To credibly argue that the family environment for both siblings is similar, I estimate the effects of maternal employment on a closely spaced sibling sample.[22] As another informal test of the assumptions of the FE model, I estimate my FE regressions separately only including women who respectively start or stop working and non-changing women with respect to employment. In the sibling sample 446 women are in the “starter” group and 559 women are in the “quitter” group. The changers who identify the effect of employment in my model should have effects of opposite sign and similar size if the decision to work/stop working early in pregnancy is exogenous conditional on my controls and a mother FE.

Table 10 shows the results for the two proposed informal tests of the validity of the sibling

[22]The share of changing mothers (with respect to employment) is the same as in the sibling sample without the spacing condition.

design. I find a smaller but consistently negative effect for maternal employment on the likelihood of premature birth in the close spacing sample and a very small and imprecisely estimated and small effect on birth weight and fetal growth. The coefficients for other controls also remain similar.

Looking separately at mothers changing their employment status in one direction only confirms the effects found for maternal employment in the main analysis. Columns three through nine present the results for the subsamples of mothers who work only in the first or the second pregnancy respectively along with non-changing mothers. Employment coefficients for starters and quitters have as expected opposite signs. For premature birth, the size of the coefficients is identically to the main result. For birth weight and fetal growth, coefficients for maternal employment are all but one insignificantly estimated and indicate at best very small effects.

Finally, to examine the role of employment using a different set of identifying assumptions, I turn to an IV strategy. I use regional unemployment rates from the 16 Danish regions for the year of birth as an instrument for maternal employment during pregnancy. Using residential information for the mother from the administrative registry data, I merge the yearly regional female unemployment rates provided by Statistics Denmark to mothers. In the cross-sectional IV model, I only use the first birth in the survey for each mother. In this framework, the IV estimator relies on variation in unemployment across regions and identifies the effect locally on mothers who change their employment due to the situation in the regional labor market. For my instrument to be valid, I assume that mothers' unobserved characteristics are uncorrelated to the level of local unemployment conditional on the set of control variables and that the unemployment rate only impacts birth outcomes through changing mothers' probability of being employed. If mothers sort into low and high unemployment regions according to their unobserved characteristics or regions vary systematically in other characteristics, my instrument is not valid. I find the assumptions made here credible in my context for two reasons: First, regions are reasonably large entities, i.e. I do not rely on municipal variation which very likely is biased by sorting of mothers. Second, regions provide a very similar level of prenatal care—very likely the most important omitted characteristic when it comes to the analysis of birth outcomes.[23]

[23]I experimented with alternative instruments using an interaction of mothers' lagged employment status (in the year before their first observed birth) and the regional unemployment rate as an instrument while

Appendix Table A.4 presents the results of my cross-sectional IV estimates. The models control for maternal age at birth, maternal cohabitation status, an indicator for white-collar occupational group, high educational group, an indicator for pregnancy complications and sick leave, and a set of year dummies. Standard errors are clustered on the regional level. As shown in the bottom of the table, the suggested instrument has good predictive power with a F-value of around 20. While the employment coefficients are very imprecisely estimated, the second stage reveals similar point estimates for maternal employment on the probability of premature birth and birth weight as found in my FE models. This finding additionally points at the validity of my FE results.

6 Conclusion

Increasing employment rates for pregnant women emphasize the importance of understanding potential effects of prenatal employment on child health. While maternal employment shortly after birth and during childhood has been studied extensively, we still lack knowledge on the causal effects of prenatal maternal employment. The analysis of prenatal maternal employment is often complicated by the lack of adequate data and unobserved differences between women, which makes a simple comparison between employed and non-employed women problematic.

In this paper, I examine the effects of prenatal maternal employment with data from Denmark. Denmark makes an important case for the study of this subject because employment rates for mothers are high by international standards both before, during and after pregnancy. Furthermore, in Denmark workplaces are obligated to accommodate tasks to the needs of pregnant women and a generous maternity leave takes pressures and stress factors off the mother's shoulders. Furthermore, financially compensated sick listing of mothers with poor health during pregnancy is an important factor in Denmark. Thus selection into and out of employment is likely different in Denmark as e.g. in the UK or the U.S.

I find that employment in the third trimester of pregnancy has no harmful effects in the Danish setting - rather than that employed women face a considerably lower risk of premature birth. As I control for sick leave from the job, I argue that this employment effect reflects an effect of actually being on the job. The risk of premature birth has been linked to maternal

including a mother FE, similar as in Currie and Cole (1993). Unfortunately, this instrument is too weak.

stress levels during pregnancy. However, mothers in Denmark are potentially exposed to more stress when not employed—resulting from anxiety over future employment prospects. Being at a workplace could in turn give women access to peers and networks and influence their behaviors during pregnancy and even the timing of their fertility decision.

The employment effect in my sample varies according to mothers' occupational group. For the risk for premature birth, the protective employment effect is bigger for blue-collar occupation mothers. This finding could suggest that common assumptions about stress and fatigue for blue-collar occupation mothers and access to "better" jobs for white-collar occupation mothers do not drive the results for premature birth. Workplaces could actually perform better in adapting tasks for blue-collar workers—and potentially oversee the need for change of tasks for white-collar mothers. However, the result could also indicate that blue-collar mothers are sick listed at a lower health threshold. Nevertheless, my findings suggest that employment should not be seen as a stress factor per se. Rather than that, maternal employment has to be studied taking countries' institutional setting—that will shape mothers' employment patterns—into account.

Additionally, I include other maternal health behaviours into the analysis to examine the importance of control for other time-varying behaviors in sibling models. I find smoking effects of similar size as earlier sibling studies. This finding indicates that these estimates are not biased by the absence of control for a wide range of control variables, which I include into the analysis. While no earlier sibling study has included maternal alcohol consumption, my findings indicate that a moderate alcohol consumption during pregnancy has harmful effects, but they are small in size.

Also the effects of maternal alcohol consumption and maternal smoking are of different size for mothers with different occupational status. This difference may be explained with beliefs about the harmfulness of the two behaviors at hand or different consumption patterns. Some evidence for Denmark suggests that mothers in the white-collar occupational groups drink more regularly also during pregnancy (see e.g., Kesmodel and Schiøler Kesmodel, 2002). However, the findings could also be due to different reporting patterns. Furthermore, as I only consider live births in this paper, my results do not take differential mortality due to fetal death into account. However, the finding calls into question common beliefs about the positive correlation of maternal health behaviors, i.e., that mothers chose healthy bundles of behaviors.

References

Abrevaya, J, "Estimating the effect of smoking on birth outcomes using a matched panel data approach," *Journal of Applied Econometrics*, 2006, *21* (4), 489–519.

_ **and CM Dahl**, "The Effects of Birth Inputs on Birthweight," *Journal of Business & Economic Statistics*, 2008, *26*, 379–397.

Aizer, A, L Stroud, and S Buka, "Maternal Stress and Child Well-Being: Evidence from Siblings," Technical Report 2009.

Almond, D and J Currie, "Chapter 15: Human Capital Development before Age Five," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4, Part 2 of *Handbook of Labor Economics*, Elsevier, 2011, pp. 1315–1486.

Baker, M and K Milligan, "Evidence from Maternity Leave Expansions of the Impact of Maternal Care on Early Child Development," *Journal of Human Resources*, 2010, *45* (1).

Black, SE, PJ Devereux, and KG Salvanes, "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *The Quarterly Journal of Economics*, 02 2007, *122* (1), 409–439.

Brooks–Gunn, J, WJ Han, and J Waldfogel, "Maternal Employment and Child Cognitive Outcomes in the First Three Years of Life: The NICHD Study of Early Child Care," *Child Development*, 2002, *73* (4), 1052–1072.

Browning, M, AM Dano, and E Heinesen, "Job displacement and stress-related health outcomes," *Health Economics*, 2006, *15* (10), 1061–1075.

Case, A, D Lubotsky, and C Paxson, "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 2002, *92* (5), 1308–1334.

Chou, SY, JT Liu, M Grossman, and TJ Joyce, "Parental Education and Child Health: Evidence from a Natural Experiment in Taiwan," Working Paper 13466, National Bureau of Economic Research October 2007.

Ciliberto, F, AR Miller, HS Nielsen, and M Simonsen, "Playing the fertility game at work," unpublished working paper 2010.

Copper, RL, RL Goldenberg, A Das, N Elder, M Swain, G Norma, R Ramsey, P Cotroneo, BA Collins, F Johnson, P Jones, and A Meier, "The preterm prediction study: Maternal stress is associated with spontaneous preterm birth at less than thirty-five weeks' gestation," *American Journal of Obstetrics and Gynecology*, 1996, *175* (5), 1286 – 1292.

Cunha, F and J Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

Currie, J, "Chapter 19: Child health in developed countries," in Anthony J. Culyer and Joseph P. Newhouse, eds., , Vol. 1, Part 2 of *Handbook of Health Economics*, Elsevier, 2000, pp. 1053 – 1090.

_ **and E Moretti**, "Mother's Education And The Intergenerational Transmission Of Human Capital: Evidence From College Openings," *The Quarterly Journal of Economics*, November 2003, *118* (4), 1495–1532.

_ **and** _ , "Biology as Destiny? Short- and Long-Run Determinants of Intergenerational Transmission of Birth Weight," *Journal of Labor Economics*, 2007, *25*, 231–264.

_ **and N Cole**, "Welfare and Child Health: The Link Between AFDC Participation and Birth Weight," *The American Economic Review*, 1993, *83* (4), pp. 971–985.

_ **, M Neidell, and JF Schmieder**, "Air pollution and infant health: Lessons from New Jersey," *Journal of Health Economics*, May 2009, *28* (3), 688–703.

_ **, M Stabile, P Manivong, and LL Roos**, "Child Health and Young Adult Outcomes," *Journal of Human Resources*, 2010, *45* (3), 517–548.

Danish Ministry of Employment, "What do we know about pregnant women's sick leave? [Analyse af graviditetsbetinget fravaer.]," Technical Report 2010.

_ , "Women and Men on the Labor Market 2011 [Kvinder og Mænd på Arbejdsmarkedet 2011.]," Technical Report 2011.

Datta Gupta, N and M Simonsen, "Non-cognitive child outcomes and universal high quality child care," *Journal of Public Economics*, February 2010, *94* (1-2), 30–43.

_ , **N Smith, and M Verner**, "Child Care and Parental Leave in the Nordic Countries: A Model to Aspire to?," IZA Discussion Papers 2014, Institute for the Study of Labor (IZA) March 2006.

Dave, D, S Decker, R Kaestner, and KI Simon, "Re-examining the Effects of Medicaid Expansions for Pregnant Women," NBER Working Papers, National Bureau of Economic Research, Inc December 2008.

Del Bono, E, J Ermisch, and M Francesconi, "Intrafamily Resource Allocations: A Dynamic Model of Birth Weight," IZA Discussion Papers 3704, Institute for the Study of Labor (IZA) September 2008.

Dustmann, C and U Schönberg, "The Effect of Expansions in Maternity Leave Coverage on Children's Long-Term Outcomes," IZA Discussion Papers 3605, Institute for the Study of Labor (IZA) July 2008.

Gabbe, SG and LP Turner, "Reproductive hazards of the American lifestyle: Work during pregnancy," *American Journal of Obstetrics and Gynecology*, 1997, *176* (4), 826 – 832.

Greve, J, "New results on the effect of maternal work hours on children's overweight status: Does the quality of child care matter?," *Labour Economics*, 2011, *18* (5), 579 – 590.

Griliches, Z, "Sibling Models and Data in Economics: Beginnings of a Survey," *Journal of Political Economy*, October 1979, *87* (5), S37–64.

Guendelman, S, M Pearl, S Graham, A Hubbard, A Hosang, and M Kharrazi, "Maternity Leave In The Ninth Month of Pregnancy and Birth Outcomes Among Working Women," *Women's Health Issues*, 2009, *19* (1), 30 – 37.

Hedegaard, M, TB Henriksen, NJ Secher, MC Hatch, and S Sabroe, "Do Stressful Life Events Affect Duration of Gestation and Risk of Preterm Delivery?," *Epidemiology*, 1996, *7* (4), 339–345.

Henderson, J, R Gray, and P Brocklehurst, "Systematic review of effects of low-moderate prenatal alcohol exposure on pregnancy outcome.," *BJOG an international journal of obstetrics and gynaecology*, 2007, *114* (3), 243–252.

Hensvik, L and P Nilsson, "Businesses, buddies and babies: social ties and fertility at work," Working Paper Series 2010:9, IFAU - Institute for Labour Market Policy Evaluation June 2010.

Johnson, T, "Maternity Leave and Employment Patterns: 2001–2003.," Technical Report, Current Population Report, P70-113. U.S. Census Bureau, Washington, DC 2007.

Kesmodel, U and P Schiøler Kesmodel, "Drinking During Pregnancy: Attitudes and Knowledge Among Pregnant Danish Women, 1998," *Alcoholism: Clinical and Experimental Research*, 2002, *26* (10), 1553–1560.

Kuhn, A, R Lalive, and J Zweimüller, "The public health costs of job loss," *Journal of Health Economics*, December 2009, *28* (6), 1099–1115.

Lien, DS and WN Evans, "Estimating the Impact of Large Cigarette Tax Hikes: The Case of Maternal Smoking and Infant Birth Weight," *The Journal of Human Resources*, 2005, *40* (2), pp. 373–392.

Lindeboom, M, A Llena-Nozal, and B van der Klaauw, "Parental education and child health: Evidence from a schooling reform," *Journal of Health Economics*, 2009, *28* (1), 109 – 131.

McCrary, J and H Royer, "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth," *American Economic Review*, 2011, *101* (1), 158–95.

Nathanson, V, N Jayesinghe, and N Roycroft, "Is it all right for women to drink small amounts of alcohol in pregnancy? No," *BMJ*, 10 2007, *335* (7625), 857–857.

Nilsson, P, "Does a pint a day affect your child's pay? The effect of prenatal alcohol exposure on adult outcomes," Working Paper Series 2008:4, IFAU - Institute for Labour Market Policy Evaluation March 2008.

Nohr, E, M Frydenberg, TB Henriksen, and J Olsen, "Does Low Participation in Cohort Studies Induce Bias?," *Epidemiology*, 2006, *17* (4), 413–418.

O'Brien, P, "Is it all right for women to drink small amounts of alcohol in pregnancy? Yes," *BMJ*, 10 2007, *335* (7625), 856–856.

Olsen, J, G Frische, AO Poulsen, and H Kirchheiner, "Changing smoking, drinking, and eating behaviour among pregnant women in Denmark. Evaluation of a health campaign in a local region.," *Scandinavian Journal of Social Medicine*, 1989, *17* (4), 277–280.

Oreopoulos, P, M Stabile, R Walld, and LL Roos, "Short-, Medium-, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins," *Journal of Human Resources*, 2008, *43* (1), 88–138.

Rasmussen, AW, "Increasing the length of parents' birth-related leave: The effect on children's long-term educational outcomes," *Labour Economics*, 2010, *17* (1), 91–100.

Rosenzweig, MR and KI Wolpin, "Sisters, Siblings, and Mothers: The Effect of Teen-Age Childbearing on Birth Outcomes in a Dynamic Family Context," *Econometrica*, March 1995, *63* (2), 303–26.

_ **and TP Schultz**, "Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight," *Journal of Political Economy*, October 1983, *91* (5), 723–46.

Rossin, M, "The effects of maternity leave on children's birth and infant health outcomes in the United States," *Journal of Health Economics*, 2011, *30* (2), 221–239.

Ruhm, CJ, "Parental leave and child health," *Journal of Health Economics*, November 2000, *19* (6), 931–960.

_ , "Parental Employment and Child Cognitive Development," *Journal of Human Resources*, 2004, *39* (1), 155–192.

Smith, JP, "The Impact of Childhood Health on Adult Labor Market Outcomes," *The Review of Economics and Statistics*, 01 2009, *91* (3), 478–489.

Strandberg-Larsen, K, N Rod Nielsen, AM Nybo Andersen, J Olsen, and M Grønbæk, "Characteristics of women who binge drink before and after they become aware of their pregnancy," *European Journal of Epidemiology*, 2008, *23*, 565–572. 10.1007/s10654-008-9265-z.

Sullivan, D and T von Wachter, "Job Displacement and Mortality: An Analysis Using Administrative Data," *The Quarterly Journal of Economics*, 2009, *124* (3), 1265–1306.

Table 1: Summary statistics, means and standard deviations, 1998-2003

	Sibling sample	*Full survey sample*	*All singleton births*
	(1)	(2)	(3)
Birth weight	3645.831	3603.040	3535.200
	(522.426)	(534.770)	(581.753)
N	10194	76281	367637
Preterm birth	0.030	0.035	0.050
Gestation	40.090	40.042	39.858
	(1.511)	(1.567)	(1.858)
Employment t-1	0.887	0.885	0.801
Manager	0.015	0.016	0.019
Professional	0.202	0.172	0.147
Technicians and associate professionals	0.321	0.299	0.256
Clerical support workers	0.197	0.206	0.202
Service and sales	0.165	0.193	0.222
Skilled agriculture, forestry, fishery	0.006	0.006	0.006
Craft and related trades workers	0.021	0.025	0.027
Plant and machine operators, assemblers	0.028	0.035	0.046
Elementary occupations	0.042	0.047	0.074
White-collar occupation	0.659	0.618	0.496
Mom smoked, registry	0.113	0.167	0.214
Pregnancy complication	0.205	0.220	0.227

Continued on the next page.

Table 1 *continued.*

	Sibling sample	Full survey sample	All singleton births
Mom's age	29.565	29.917	29.520
	(3.822)	(4.264)	(4.764)
Low education mom	0.090	0.128	0.225
High education mom	0.385	0.330	0.261
Parity	1.775	1.726	1.842
	(0.779)	(0.824)	(0.957)
Sex (female)	0.487	0.488	0.487
Firstborn	0.395	0.468	0.430
Mom cohabiting	0.942	0.910	0.874
Employment, int 1	0.814	0.807	
Sick listed, int 1	0.060	0.067	
Employment, int 2	0.772	0.772	
Sick listed, int 2	0.103	0.120	
Mom smoked, int 1	0.194	0.256	
Mom smoked, int 2	0.113	0.163	
Mom cunsumed alcohol, int 1	0.260	0.244	
Mom consumed alcohol, int 2	0.326	0.310	
Mom excercised, int 2	0.316	0.305	
Number of times fish/week, int 1	2.230	2.187	
	(1.855)	(1.904)	
BMI, prepregnancy	23.514	23.584	
	(4.197)	(4.248)	
BMI, int 2	26.955	27.112	
	(4.010)	(4.067)	

Continued on the next page.

Table 1 *continued.*

	Sibling sample	Full survey sample	All singleton births
Number of cigs, int 1	7.603	8.249	
	(5.615)	(5.642)	
N	1981	19486	
Number of cigs, int 2	8.513	8.667	
	(5.864)	(5.367)	
N	1152	12449	
Number of glasses/week	1.737	1.824	
	(1.119)	(1.272)	
N	2652	18575	
Number of glasses/week, int2	1.889	1.930	
	(1.301)	(1.331)	
N	3325	23661	
Hours worked, int 1	35.554	35.396	
	(5.275)	(5.805)	
Hours worked int 1 of working moms at int 2	35.205	35.127	
	(6.396)	(6.708)	

Table 2: Summary Statistics for changing and non-changing mothers according to employment status

	Starters		*Quitters*		*Always working*		*Never working*	
	First	*Second*	*First*	*Second*	*First*	*Second*	*First*	*Second*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth weight	3602.973	3737.711	3583.751	3679.247	3576.573	3715.154	3585.014	3667.005
	(490.414)	(457.706)	(543.236)	(537.610)	(530.841)	(503.527)	(542.394)	(526.419)
Preterm birth	0.043	0.007	0.038	0.050	0.035	0.021	0.034	0.030
Fetal growth	90.478	93.907	90.144	93.180	89.776	93.595	90.115	92.777
	(11.123)	(10.845)	(12.511)	(12.318)	(12.165)	(11.833)	(12.432)	(12.355)
Sick listed, int 2	-	0.141	0.272	-	0.125	0.120	-	-
Pregnancy complication	0.247	0.188	0.256	0.166	0.232	0.177	0.201	0.183
Mom smoked, int 2	0.146	0.137	0.168	0.148	0.103	0.085	0.175	0.158
Mom consumed alcohol, int 2	0.303	0.342	0.252	0.284	0.333	0.365	0.252	0.266
Mom cohabiting	0.861	0.984	0.864	0.961	0.922	0.990	0.843	0.956
Mom's age	26.791	29.466	27.970	30.009	28.875	31.288	27.148	29.402
	(3.577)	(3.577)	(3.980)	(3.969)	(3.402)	(3.387)	(3.893)	(3.982)
High education mom	0.216	0.508	0.235	0.253	0.440	0.464	0.122	0.215
White-collar occupation	0.516	0.659	0.478	0.506	0.752	0.752	0.338	0.405
N	446	446	559	559	3467	3393	656	640

Notes: Std.dev. in parentheses; first and second observed births for mothers in the sibling sample.

Table 3: OLS estimates for the effect of maternal inputs on premature birth, birth weight and fetal growth

	Sibling sample				*Full survey sample*			
	Premature		*Birth weight*		*Premature*		*Birth weight*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Employment, int 2	-0.010**	-0.014***	0.154	-1.414	-0.005***	-0.008***	4.286	4.714
	(0.004)	(0.005)	(13.595)	(13.927)	(0.002)	(0.002)	(4.721)	(4.889)
White-collar occu.		0.005		3.732		0.000		1.748
		(0.006)		(19.076)		(0.002)		(6.756)
Blue-collar occu.		0.011		28.819		0.002		3.876
		(0.007)		(19.884)		(0.003)		(6.887)
Sick listed, int 2		0.019***		5.480		0.017***		-4.405
		(0.007)		(17.434)		(0.002)		(6.190)
Pregnancy compl.		0.034***		-33.282**		0.049***		-71.910***
		(0.005)		(14.699)		(0.002)		(5.186)
Mom smoked, int 2	0.019***	0.011	-231.970***	-116.098***	0.007***	-0.002	-240.781***	-117.840***
	(0.007)	(0.011)	(19.298)	(31.689)	(0.002)	(0.003)	(5.392)	(9.141)
Number of cigs, int 2		0.001		-14.067***		0.001**		-13.975***
		(0.001)		(3.226)		(0.000)		(0.891)
Glasses alc./week, int 2		0.001		-11.655**		-0.001		-4.558***
		(0.002)		(4.933)		(0.001)		(1.640)
N	10194	10194	10194	10194	76281	76281	76281	76281
R^2	0.005	0.014	0.062	0.081	0.004	0.018	0.073	0.090

Notes: Additional controls in the narrow control model: maternal educational level, maternal age, cohabitation status at birth, child sex and whether the child is the first-born, set of year dummies, an indicator for maternal employment and an indicator for smoking at interview two. Additional controls in the extended control model: an indicator for maternal sick leave during pregnancy (more than three days), an indicator for pregnancy-related health problems, indicators for the mother's occupational group (omitted category is missing occupation in t-1), whether the mother smoked, whether she exercised and consumed fish, the number of daily cigarettes, the units of alcohol per week and the pre-pregnancy BMI; clustered std. errors in parentheses; ***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table 4: FE estimates for the effect of maternal inputs on premature birth, birth weight and fetal growth, narrow set of controls

	Premature	*Birth weight*	*Fetal growth*
	(1)	(2)	(3)
Employment, int 2	-0.027***	16.956	0.078
	(0.008)	(16.381)	(0.370)
Mom smoked, int 2	-0.000	-95.900**	-2.397***
	(0.015)	(37.557)	(0.851)
N	10194	10194	10194
R^2	0.010	0.094	0.126

Notes: For control variables see Notes to Table 3; clustered std. errors in parentheses; ***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table 5: FE estimates for the effect of maternal inputs on premature birth, birth weight and fetal growth, extended set of controls

	Premature	*Birth weight*	*Fetal growth*
	(1)	(2)	(3)
Employment, int 2	-0.028***	11.436	-0.032
	(0.008)	(17.088)	(0.387)
White-collar occu.	-0.003	46.206*	0.957*
	(0.010)	(25.547)	(0.573)
Blue-collar occu.	-0.016	16.641	0.071
	(0.012)	(26.370)	(0.594)
Sick listed, int 2	0.007	28.478	0.727
	(0.010)	(20.660)	(0.459)
Pregnancy compl.	0.025***	-57.128***	-1.747***
	(0.007)	(15.838)	(0.356)
Mom smoked, int 2	-0.000	-94.489**	-2.364***
	(0.015)	(37.594)	(0.849)
Glasses alc./week, int2	0.008***	-18.080**	-0.385**
	(0.003)	(7.070)	(0.159)
N	10194	10194	10194
R^2	0.016	0.100	0.134

Notes: For control variables see Notes to Table 3 with the exception of number of daily cigarettes; clustered std. errors in parentheses;***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table 6: FE estimates for the effect of maternal inputs on premature birth and birth weight by educational group and occupational group

	High education		*Low education*		*White-collar occup. group*		*Blue-collar occup. group*	
	Premature	*Birth weight*	*Premature*	*Birth weight*	*Premature*	*Birth weight*	*Premature*	*Birth weight*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Employment, int 2	-0.033**	21.440	-0.021**	-2.099	-0.032***	48.974**	-0.058***	15.854
	(0.014)	(31.877)	(0.009)	(19.607)	(0.011)	(23.749)	(0.015)	(29.955)
White-collar occu.	0.014	54.291	-0.009	40.182				
	(0.014)	(45.990)	(0.012)	(29.478)				
Blue-collar occu.	-0.027	38.869	-0.016	20.454				
	(0.030)	(69.784)	(0.013)	(28.740)				
Sick listed, int 2	0.023	32.761	-0.003	27.349	0.005	27.373	0.018	16.957
	(0.017)	(35.782)	(0.013)	(25.139)	(0.013)	(28.679)	(0.018)	(33.075)
Pregnancy compl.	0.031***	-47.905*	0.021**	-64.974***	0.036***	-67.164***	0.018	-35.603
	(0.010)	(24.502)	(0.009)	(20.826)	(0.008)	(19.330)	(0.015)	(31.710)
Mom smoked, int 2	-0.010*	-33.787	0.002	-121.170***	-0.026	-35.091	0.032	-168.042***
	(0.006)	(63.114)	(0.021)	(46.184)	(0.018)	(50.272)	(0.022)	(64.271)
Glasses alc./week, int 2	0.011**	-27.932**	0.006	-10.321	0.009**	-14.522*	0.004	-22.248
	(0.005)	(11.246)	(0.004)	(8.954)	(0.004)	(8.495)	(0.007)	(15.030)
N	4226	4226	5949	5949	6591	6591	2612	2612
R^2	0.021	0.082	0.014	0.116	0.021	0.097	0.026	0.120

Notes: For control variables see Notes to Table 3 with the exception of number of daily cigarettes; clustered std. errors in parentheses;***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table 7: Probability of very premature birth (<30 weeks of gestation) for survey mothers and all mothers

	Very premature birth					
Nr. of siblings in survey	0		1		Total	
	Obs	%	Obs	%	Obs	%
1	73,749	99.7	226	0.3	73,975	100
2	10,788	99.8	24	0.2	10,812	100
3	180	100	0	0	180	100
Not in survey	281,484	99.6	1,186	0.4	282,670	100
Total	366,201	99.6	1,436	0.4	367,637.0	100.0

Notes: Survey children are those for whom mothers have at least answered the first pregnancy interview around week 12. The outcome measure comes from the administrative register data.

Table 8: Robustness: Probability of being pregnant or missing at interview 2

	Pregnant at interview 2	*Missing interview 2*
	(1)	(2)
Timing of interview	0.548***	
	(0.080)	
Employment, int 1	-0.003	-0.017
	(0.003)	(0.011)
Sick listed, int 1	-0.004	0.021
	(0.006)	(0.013)
Pregnancy complication	-0.004	0.017**
	(0.003)	(0.008)
White-collar occupation	0.007*	0.017
	(0.004)	(0.014)
Blue-collar occupation	-0.002	0.001
	(0.005)	(0.014)
Mom smoked, int 1	-0.001	-0.004
	(0.004)	(0.013)
Number of glasses/week	-0.001	0.001
	(0.002)	(0.005)
N	10306	10845
Mother FE	yes	yes

Notes: Additional control variables: child sex and indicator for first parity, mothers' educational group and age, mother exercised, consumed fish and prepregnancy BMI, year dummies; clustered std. errors in parentheses; ***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table 9: Robustness: Effect of employment and other health inputs at interview one and imputed employment on probability of premature birth

	Very premature	*Premature*	*Very premature*	*Premature*		*Premature*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Employment, int 1	-0.001	-0.007***	0.001	-0.018*			
	(0.001)	(0.002)	(0.002)	(0.009)			
Employment, int 2					-0.026***		
					(0.008)		
Employment imputed (1)						-0.015*	
						(0.008)	
Employment imputed (0)							-0.043***
							(0.007)
Sick listed, int 1	0.002**	0.022***	-0.004	-0.007	-0.010	-0.008	-0.006
	(0.001)	(0.003)	(0.004)	(0.012)	(0.013)	(0.012)	(0.012)
Pregnancy complication	0.005***	0.064***	0.003*	0.041***	0.026***	0.041***	0.041***
	(0.001)	(0.002)	(0.002)	(0.007)	(0.007)	(0.007)	(0.007)
White-collar occupation	-0.001	-0.001	-0.001	-0.006	-0.004	-0.007	-0.005
	(0.001)	(0.003)	(0.003)	(0.010)	(0.010)	(0.010)	(0.010)
Blue-collar occupation	-0.001	0.001	0.000	-0.019	-0.015	-0.019	-0.016
	(0.001)	(0.003)	(0.003)	(0.012)	(0.012)	(0.012)	(0.012)
Mom smoked, int 1	0.002***	0.006***	-0.001	-0.013	-0.015	-0.013	-0.012
	(0.001)	(0.002)	(0.001)	(0.010)	(0.011)	(0.010)	(0.010)
Number of glasses/week	-0.000	-0.001	0.001	0.004	0.003	0.004	0.003
	(0.000)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
N	83480	83480	10845	10845	10199	10847	10847
Mother FE	no	no	yes	yes	yes	yes	yes

Notes: Additional control variables see Notes to Table 8; clustered std. errors in parentheses; ***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table 10: Robustness: Closely spaced siblings and FE for starters/quitters

	Sibling sample, close spacing			*Starters*	*Quitters*	*Starters*	*Quitters*	*Starters*	*Quitters*
	Premature	*Birth weight*	*Fetal Grouth*	*Premature*		*Birth weight*		*Fetal Grouth*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Employment, int 2	-0.017*	-8.259	-0.313						
	(0.009)	(21.977)	(0.509)						
Mom starts				-0.029**		12.442		0.020	
				(0.013)		(26.436)		(0.594)	
Mom stops					0.025**		-4.187		0.190
					(0.011)		(24.247)		(0.552)
White-collar occu.	-0.022*	96.212***	1.809**	-0.000	-0.009	35.768	60.410**	0.734	1.218*
	(0.013)	(36.045)	(0.816)	(0.010)	(0.011)	(27.537)	(28.793)	(0.626)	(0.643)
Blue-collar occu.	-0.036**	74.267**	1.412*	-0.016	-0.019	32.223	26.999	0.474	0.167
	(0.016)	(36.449)	(0.830)	(0.013)	(0.014)	(29.238)	(29.844)	(0.661)	(0.671)
Sick listed, int 2	0.013	15.721	0.395	0.001	0.010	47.529**	26.733	1.175**	0.671
	(0.012)	(27.057)	(0.612)	(0.011)	(0.011)	(22.954)	(21.982)	(0.511)	(0.489)
Pregnancy compl.	0.026***	-54.978**	-1.747***	0.028***	0.027***	-56.906***	-61.904***	-1.745***	-1.819***
	(0.009)	(21.853)	(0.494)	(0.007)	(0.007)	(16.743)	(16.857)	(0.376)	(0.378)
Mom smoked, int 2	0.003	-75.148	-1.917	-0.012	0.006	-91.527**	-91.432**	-2.396***	-2.239**
	(0.015)	(51.543)	(1.189)	(0.017)	(0.017)	(40.927)	(40.422)	(0.922)	(0.912)
Glasses alc./week, int 2	0.010**	-12.568	-0.178	0.007**	0.008**	-24.846***	-20.915***	-0.551***	-0.424**
	(0.005)	(9.833)	(0.222)	(0.003)	(0.003)	(7.528)	(7.348)	(0.170)	(0.167)
N	5829	5829	5829	8966	9197	8966	9197	8966	9197
R^2	0.021	0.100	0.132	0.018	0.014	0.106	0.102	0.140	0.136

Notes: For control variables see Notes to Table 3 except for number of daily cigs; clustered std. errors in parentheses;***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

A Appendix

Table A.1: Logistic regression for second pregnancy behavior (work) on first pregnancy behaviors and outcomes

	Work, pregnany 2	
	(1)	(2)
Low birth weight	1.173	
	(0.334)	
Preterm birth		1.418
		(0.320)
Employment, int 2	6.143***	6.165***
	(0.551)	(0.553)
White-collar occu.	2.160***	2.164***
	(0.268)	(0.269)
Blue-collar occu.	1.160	1.159
	(0.147)	(0.147)
Sick-listed, int 2	0.496***	0.491***
	(0.057)	(0.057)
Pregnancy complications	0.981	0.975
	(0.091)	(0.091)
Mom smoked, int 2	1.103	1.098
	(0.212)	(0.211)
N	4700	4700

Notes: Odds Ratios; std. errors in parentheses; Additional controls: mothers' educational level, child's sex, mother's age, mother's cohabitation status, number of cigs and units of alcohol, mother excercises, pre-pregnancy BMI, fish consumption dummy; all controls for behaviors in first observed pregnancy; ***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table A.2: OLS estimates for the effect of maternal inputs on fetal growth

	Sibling sample		*Full survey sample*	
	(1)	(2)	(3)	(4)
Employment, int 2	-0.222	-0.311	-0.007	-0.084
	(0.314)	(0.321)	(0.108)	(0.112)
White-collar occupation		0.130		0.121
		(0.441)		(0.155)
Blue-collar occupation		0.621		0.122
		(0.459)		(0.158)
Sick listed, int 2		0.484		0.320**
		(0.397)		(0.141)
Pregnancy complication		-1.250***		-1.944***
		(0.328)		(0.116)
Mom smoked, int 2	-5.466***	-2.922***	-5.791***	-3.028***
	(0.440)	(0.725)	(0.124)	(0.210)
Number of cigs, int 2		-0.310***		-0.315***
		(0.073)		(0.021)
Number of glasses/week, int2		-0.323***		-0.169***
		(0.111)		(0.038)
N	10194	10194	76281	76281
R^2	0.076	0.096	0.085	0.103

*Notes:*clustered std. errors in parentheses; controls: see Table 3; ***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table A.3: FE estimates for the effect of maternal inputs on fetal growth by educational and occupational group

	High edu.	*Low edu.*	*White-collar*	*Blue-collar*
	(1)	(2)	(3)	(4)
Employment, int 2	0.345	-0.363	0.804	-0.209
	(0.751)	(0.438)	(0.552)	(0.664)
White-collar occu.	1.322	0.741		
	(1.049)	(0.661)		
Blue-collar occu.	0.368	0.163		
	(1.567)	(0.646)		
Sick listed, int 2	1.054	0.581	0.644	0.628
	(0.787)	(0.562)	(0.636)	(0.741)
Pregnancy compl.	-1.529***	-1.938***	-1.928***	-1.233*
	(0.554)	(0.465)	(0.441)	(0.692)
Mom smoked, int 2	-1.568	-2.758***	-1.135	-4.161***
	(1.558)	(1.019)	(1.149)	(1.490)
Glasses alc./week, int 2	-0.562**	-0.241	-0.313	-0.443
	(0.249)	(0.206)	(0.193)	(0.337)
N	4226	5949	6591	2612
R^2	0.118	0.148	0.131	0.157

Notes: Clustered std. errors in parentheses; controls: see Table 5; ***significant at the 1 percent level, **significant at the 5 percent level *significant at the 10 percent level

Table A.4: Robustness: IV estimates for the effect of maternal employment on premature birth and birth weight

	Premature birth	*Birth weight*
	(1)	(2)
Employment, int 2	-0.049	94.497
	(0.045)	(180.632)
White-collar occupation	0.013	-10.229
	(0.015)	(61.674)
Blue-colalr occupation	0.010	-12.231
	(0.010)	(41.674)
Pregnancy complication	0.051***	-76.333***
	(0.003)	(7.751)
N	73421	73421
R^2	0.008	0.020
First stage F value for instrument	*20*	

Notes: Instrument: Regional female unemployment rate at year t; sample includes first observed birth per mother; clustered std. errors (at regional level) in parentheses; Additional controls: child sex, mother's age, mother's cohabitation status, dummy for high educational level of mother, year dummies; ***significant at the 1 percent level

Early Interventions and Health Outcomes: Evidence from the Danish Home Visiting Program

Miriam Wüst[*]

Abstract

This paper uses unique historical data from Denmark to estimate the effects of the introduction of a universal home visiting program for mothers and their infants from 1937 through 1949. To identify the effects of the program on infant first year survival rates, the paper exploits exogenous variation in the timing of program implementation across municipalities. Using data for all Danish towns, I find a significant and positive effect on log infant survival rates of around 0.5-0.8 percent or around 5-8 lives saved per 1000 live births at the mean infant survival rate for the period under consideration. I find that the effect of the program was strongest in the great majority of relatively small Danish towns of the time—most likely due to pre-existing and confounding efforts to fight infant mortality in the five major metropolitan areas. Consistent with this finding, my result for the small town sample is robust to the inclusion of time trends, as well as to the inclusion of time-varying controls for socio-economic town characteristics. Using complementary data at the medical district level for all urban and rural areas in Denmark and adding controls for other characteristics of the medical system of the time, I confirm that the program had a positive effect on infant survival. Moreover, testing for potential mechanisms, I find that the program decreased mortality rates for acute enteritis, a major cause of infant death at the time. This finding indicates that one important mechanism for the effect of the home visiting program was the nurses' promotion of breastfeeding and proper infant nutrition. A stylized analysis of the costs of the home visiting program shows that those were modest

[*]I thank Nabanita Datta Gupta, Tor Eriksson, Paul Bingley, Janet Currie, two anonymous referees and seminar participants at SFI and Aarhus University, as well as the EALE 2011 conference, for helpful comments. Parts of this paper were written during my stay at Columbia University, NYC, and I thank fellow graduate students for helpful discussions. I also thank the Danish Twin Registry for allowing me to use their data and Axel Skytthe for valuable help with the data. Kirsten Krøyer (SFI), Peder Gammeltoft and colleagues from the "Digital atlas of the Danish historical-administrative geography"-project, the Danish Data Archive and the Danish National Archives provided valuable help in finding data for this project. I acknowledge financial support by the Danish Agency for Science, Technology and Innovation through a grant to the Graduate School for Integration, Production and Welfare.

when, e.g., compared to another large scale public health intervention of the time, clean water supply, and home visiting in the U.S. Finally, taking a first glance at longer-run adult outcomes shown to correlate with early life nutrition, my preliminary findings show that treated individuals are shorter and more likely to be overweight in early adulthood. While these findings are consistent with a scarring story, they are based on a sample of twins and should therefore be interpreted with caution. Nevertheless, they could indicate that the home visiting program is a potential environmental factor explaining a sharp increase in obesity rates in Denmark for the 1940s birth cohorts. Future research should use more comprehensive data to examine this factor further, as well as including other long-run outcomes.

1 Introduction

A growing body of evidence suggests that early life health conditions have long-lasting consequences for adult outcomes, among them educational attainment, income, and health (Almond and Currie, 2011; Ben-Shlomo and Kuh, 2002). Studies both in the economics and epidemiology literature find an important role for very early life conditions—such as adverse conditions in utero or high mortality rates in infancy—for explaining inequalities in health and socio-economic status in adulthood (Forsdahl, 1979; Barker, ed, 1992; Almond, 2006; van den Berg et al., 2006).

Economists take their point of departure from models of human capital and health capital formation. These models emphasize critical investments in infants, investments that are likely to be complements to later ones (Cunha and Heckman, 2007). Thus one important role for early interventions is securing and encouraging investments for which there are no substitutes. The Danish home visiting program for mothers and infants was created as such an intervention: It focused on the promotion of parental investments in infants, most importantly breastfeeding, a factor that had earlier been identified as crucial to decreasing infant mortality. Exploiting exogenous variation in program roll-out, this paper is the first to examine the causal effects of the introduction of a universal home visiting program. Although the program has changed considerably over time, it remains in place today as a universal offer to all infants and their mothers.

While this program could be seen as a feature specific to Denmark, it has characteristics that make its analysis applicable to both the early influence and development economics literature. This paper makes three contributions to these strands of literature: First, it focuses on the early life origins of health disparities and the impact of related policy. While a number of recent studies focus on the impact of general health conditions around birth and infancy for later life outcomes (Almond, 2006; van den Berg et al., 2006; Bozzoli et al., 2007; Delaney et al., 2009), few evaluate *early interventions* that can cushion the long-run consequences of health shocks (for important exceptions see Cutler and Miller, 2005; Almond and Chay, 2006; Bhalotra and Venkataramani, 2011). As this paper evaluates the importance of an early intervention, it has more clearly defined policy implications.

Second, as home visiting programs may be useful in a number of developing countries today (Engle et al., 2007), knowledge about the effects of these programs is in demand in development economics. However, the lack of data and adequate identification strategies often complicates program evaluation. Denmark in the 1930s was a high-infant mortality country struggling with infectious disease as one main cause of death in the first year of life (DNBH, 1970). Thus a look at historical evidence from Denmark can be useful for informing policies in similar contexts today. Third, while a number of studies have examined the effects of very targeted home visiting programs in the U.S. (for an overview see table 9 in Almond and Currie, 2011), we lack knowledge about the potential benefits of universal home visiting (Kamerman and Kahn, 1993). This paper provides this kind of evidence.

To my best knowledge, only one other paper has examined the impact of home visiting on infant mortality in a similar context: Moehling and Thomasson (2012b) use data on U.S. state spending on different activities under the 1920s "Shepard-Towner" Act. This federal public health initiative aimed at improving maternal and infant health by supporting programs such as health conferences and home visiting (see also Moehling and Thomasson, 2012a). They conclude that "direct and personal interventions such as nurse visits were most effective at reducing infant mortality" (Moehling and Thomasson, 2012b, p.6) and that activities under the Shepard-Towner Act accounted for around 9-21 percent of the mortality decline of the period. Moreover, while not studying specific public health interventions, Miller (2008) refers to programs such as home visiting as a major mechanism underlying the effect of the expansion of U.S. women's suffrage on the child mortality decline. Interestingly, he finds that increased public health spending decreased overall child mortality for all age groups except the under-one-year-olds. When examining specific causes of death, he finds that this child mortality decrease was driven by reductions in deaths from infectious diseases, among them diarrhea in the under-two age group.

By exploiting credibly exogenous variation in the roll-out of the Danish home visiting program, this paper provides additional evidence for the role of home visiting in explaining the infant mortality decline in the first part of the 20th century. While centrally designed and funded, the Danish program was implemented at the municipal level. A shortage of qualified nurses, a lengthy procedure for the approval of both new nurses and municipal programs by the Danish National Board of Health (DNBH), and local preferences by key actors were important factors determining the roll-out of the program across Denmark.

While the home visiting program constituted the first large-scale public health intervention for infants and new mothers in Denmark, prior to the program several local philanthropic organizations had been active in infant care. The most important initiative was infant wards for poor mothers and their infants, predominantly in the major towns. Additionally, the five major Danish towns of the period—Copenhagen/Frederiksberg,[1] Aarhus, Odense, Aalborg and Esbjerg—departed from most Danish towns with respect to population density, number of live births, and socio-economic characteristics. If these major towns had had a greater focus on health issues before the home visiting program, they potentially were more likely both to have higher infant survival rates and to quickly implement the home visiting program. As I find some indication for this suggestion in the data, I examine the robustness of my results by excluding major towns and the few small towns with infant wards.

Using data for the 87 Danish towns in 1933-1949, I find that the program significantly increased infant first-year survival rates.[2] The log infant survival rate per 1000 live births increased by around 0.005-0.008 in the year of program onset. When compared to the period's overall increase of the infant survival rate in Danish towns (roughly 4 percent from around 930 infants to around 970 infants per 1000 live births), this effect corresponds to the home visiting program's accounting for 13-20 percent of that overall increase. Looking at the timing of treatment effects, I additionally find that the effect of the program emerged only after actual treatment initiation.

My results also indicate that the program was more effective in the majority of small towns, which were much less likely to have seen preceding interventions. This finding can be interpreted along the same lines as the race-related differences that Moehling and Thomasson (2012b) find in their analysis of home visiting in the U.S.: Danish mothers in the small towns were likely to be unexposed to earlier treatments and thus had the greatest potential for improving their children's outcomes through early contact with health professionals, an increased focus on proper infant care, thereby avoiding preventable diseases.

I examine several threats to the identification strategy of this paper: I show that—as also referred to in anecdotal evidence (Buus, 2001)—pre-treatment infant mortality is a weak predictor for early treatment initiation once the major towns are omitted from the sample. My main result is robust across specifications that include year and town fixed effects and

[1] While Copenhagen and Frederiksberg are two separate administrative entities, they form one town in terms of geography. Thus I refer to them as one town in this paper.

[2] Most towns are very small and see very few births and often zero infant deaths per year. Thus I use the infant survival rate—which is close to one—in its log-transformed form as my outcome measure.

town-specific time trends. To further examine the impact of other time-varying municipal characteristics that could predict both the timing of program implementation and infant survival rates, I add socio-economic controls—the percentage of social-democratic voters and the percentage of income tax payers—to the analysis. While my estimates are similar but insignificantly estimated in my full sample, estimates that omit major towns (as well as towns with infant wards) remain similar in significance and size.

To complement my main analysis on town data and shed light on potential mechanisms, I use data for all aggregated urban and rural municipalities in the Danish medical districts. In the following I call this level of aggregation "subdistrict". While the aggregation makes it less suited for estimating the effect of the treatment, this data also contains rural areas and information on a set of specific causes of death.

In the subdistrict data, I find program effects on log infant survival of 0.004. Importantly, using medical district data, which allows for controlling for other characteristics of the medical system—the number of physicians, midwives and hospital beds available, I find significant program effects for deaths from acute enteritis. This condition is the only infant-specific cause of death, which I can examine. It accounted for a considerable percentage of infant deaths at the time and was closely related to improper preparation of infants' food.

My analysis highlights the importance of the home visiting program for the overall increase of infant survival rates in Denmark in the early 20th century—most likely through its focus on proper infant nutrition and breastfeeding, and its timely referral of sick infants to other medical professionals. Contrasting the immediate effect of the program with its costs and comparing my estimates to the ones obtained in similar studies (Cutler and Miller, 2005; Moehling and Thomasson, 2012b), I conclude that the program was highly cost-effective. I calculate that the program cost was around 159 USD per person-year saved or around 10,314 USD per life saved (in 2010 USD). Although based on a number of simplifying assumptions, these figures indicate that the program was successful at modest costs.

Exposure to lower infant mortality rates could result in long-run benefits for treated cohorts. Consequently, this paper takes a first glance at longer-run returns to the program. I focus on health outcomes that have earlier been shown to correlate with infant and childhood nutrition status. An important constraint of my analysis of long-run outcomes is data. I use military service examination records for a sample of Danish twins.[3] Therefore, the results should be interpreted with caution, and future analyses should consult encompassing population data—from the Danish administrative registers and military examination records for a broader population—to further examine the long-run effects of the home visiting program. My preliminary analyses suggest a negative effect of the program for male adults' height and the risk of overweight in early adulthood for treated men. This finding is the first empirical result supporting hypotheses about the role of the home visiting program in the development of the obesity epidemic in Denmark.

The rest of the paper unfolds as follows: Section 2 describes the necessary background, institutional features of the home visiting program, and the roll-out of the program. Section 3 describes the data and gives descriptive statistics. Section 4 presents the empirical strategy of the paper. Section 5 presents results for infant mortality and survival rates and robustness tests. Section 6 presents simple calculations for the cost-effectiveness of the program, and section 7 presents first and preliminary results for the program's long-run effects. Section 8 both discusses the results and concludes.

2 Background

2.1 The Danish Medical System and Administrative Structure of the 1930s and 1940s

During the period under consideration, Denmark consisted of 23 medical districts outside Copenhagen (see appendix figure A.1 for a map of the medical districts). In 1937, Denmark comprised 87 towns (urban municipalities) and approximately 1,200 rural municipalities (Frandsen, 1984; DNBH, 1933-1950). Although constituting a small percentage of all municipalities, the towns were already playing an important role in terms of population. Urban areas held 40 percent of the population in the 1920s, and this percentage increased during the 1930s due to social and economic changes.

Table 1 shows the average number of live births and the average size of population for two collapsed categories—the five major metropolitan areas (Copenhagen/Frederiksberg, Aarhus, Odense, Aalborg and Esbjerg) and the great majority of small and medium-sized Danish

[3] I do not exploit the twin dimension of the data.

towns. The table illustrates the special status of the major towns, which saw on average more live births and held a greater percentage of the population than all the other towns combined. Likewise, the metropolitan areas differed from the rest of the country on many other relevant dimensions, such as general housing conditions or their inhabitants' occupation and status with respect to tax liabilities, as Table 2 illustrates.

Important for the analysis in this paper—and given that the major towns had struggled with the "urban penalty," i.e., high levels of infant mortality, throughout the 19th century (Løkke, 1998)—privately organized health initiatives had existed, predominantly in these major towns before the introduction of the home visiting program. The most important one was infant wards, which provided counselling and health checks for infants. Several had already opened in the early 1900s, mainly in the Copenhagen area, and by the 1920s and 1930s a few wards had also opened in other towns (DNBH, various years). The wards served mainly poor women, and only breastfeeding women were welcomed (Løkke, 2008).[4] In the 1910s and 1920s, around 20 percent of all infants in Copenhagen received treatment in these wards (Løkke, 1998).[5] However, postnatal care for the general population of mothers was rather poor in Denmark, whereas countries such as the U.S., the UK and the Netherlands had established universal programs with home visitors (Kamerman and Kahn, 1993; Loudon, 1992; Moehling and Thomasson, 2012b).

Danish public health insurance started in 1892 (Strandberg-Larsen et al., 2007). By the 1930s around half of all Danes were covered (ibid.), and public hospitals were under medical district supervision. Half of all doctors in the early 1930s were general practitioners (GPs), evenly distributed across the country due to contracts with the health insurance program. Thus Danish mothers and infants in principle had good access to doctors. Midwives were also evenly distributed across medical districts. Danish midwifery was of high quality, with midwives usually performing births without the assistance of doctors (Loudon, 1992). Most likely as a consequence of good access to GPs and well-educated midwives, home births remained the norm in Denmark, especially in rural areas and the majority of towns, until the

[4] One important service was the distribution of milk to breastfeeding mothers as complementary nutrition. The wards campaigned for breastfeeding and thus showed a similar focus to that of the home visiting program.

[5] Data on the users of infant wards illustrates that infant mortality was relatively low, most likely due to the timing of mothers' first visit to the wards, typically after the first month of an infant's life. Thus lower mortality for treated infants resulted from a select group, i.e., infants selected on survival until month one. Infant wards—in towns where they existed—became a part of the new home visiting program. The existence of these wards in Copenhagen facilitated the implementation of the home visiting program in 1937.

1960s (Vallgårda, 1996). Again, the major towns were early movers with respect to hospital births, as illustrated by the example of Copenhagen, where the first maternity hospital opened in the late 1800s, primarily to care for unmarried and less privileged mothers.

2.2 Infant Mortality and the Emergence of the Home Visiting Program

Infant mortality in Denmark was around 11 percent in the first decade of the 20th century and by the mid-1930s was stagnating at around 6.5 percent (DNBH, various years). This percentage, however, was higher than that for neighboring countries and thus a concern for Danish medical professionals and policy makers. Figure 1 shows the mean of the log infant mortality rate for 1933-1949 based on medical district data.[6] The figure shows a clear downwards trend after 1937 from a rather high and stable level in the early 1930s.

The small size of many municipalities meant that most statistical materials of the time were not published for all municipalities. Thus Figure 2 plots the mean of the log infant mortality rates for the urban and rural subdistricts in the medical districts. Overall, urban and rural areas followed a similar trend over the period. While the urban mortality rate was below the rural rate for most of the period, the graph shows an urban “war penalty” around 1945. The uptick around 1945 for urban areas most likely results from poor economic conditions related to the end of World War II and of Denmark’s occupation by German forces (1940-1945). Although Denmark was not engaged in severe armed conflict, the constraints in these years affected urban areas the worst.

Many factors—such as economic and social changes and the introduction of new drugs, vaccines, and numerous public health initiatives—must have had an impact on the decrease of mortality over the 1930s and 1940s. (Section 5 discusses some of these topics in greater detail.) In their effort to fight high infant mortality, medical professionals had identified both the home environment and mothers’ care-taking in the initial period of their infant’s life as two important factors, especially as infectious diseases still played a major role for first-year mortality.

Appendix Table A.1 illustrates morbidity and mortality from acute enteritis for 1921-1949. As the table shows, acute enteritis was especially dangerous for infants, who accounted for the majority of deaths. In the 1930s mortality rate from acute enteritis was at 5.2 infants

[6] All figures and tables exclude the districts of Aabenraa and Sønderborg. They merge in the period, making the data for before and after the merge difficult to compare.

per 1000 live births, which roughly corresponds to 10 percent of the overall infant mortality in that decade. As the condition was often caused by poor treatment of food, the DNBH focused on the need for promoting breastfeeding.[7]

Experimenting with home visiting, the DNBH in 1929 introduced a 5-year trial with visiting nurses in three treatment and three control areas (DNBH, 1970). The U.S.-based Rockefeller Foundation financed this trial, which was supposed to establish the effect of a universal home visiting program. The trial was not based on randomization, nor did the DNBH register pre-treatment outcomes in control areas. Thus the conclusions were based on a comparison of post-treatment means for deliberately chosen treatment and control areas. Anecdotal evidence suggests that treatment and control areas were chosen for producing positive effects (Buus, 2001).[8]

Although a number of factors cast doubt on the positive results found in the trial, the DNBH recommended the expansion of the home visiting program to the entire country. The DNBH highlighted the high compliance rates among mothers as one of the most positive findings of the trial. Although the DNBH had expected compliance rates of around 60 percent, nearly all mothers welcomed the visiting nurses. In March 1937 the Danish parliament enacted the Act on the Home Visiting Program.

The DNBH designed the program with great care: Home visitors were to visit new mothers and their children in the first year of life according to an organized, uniform schedule. This schedule comprised at least 10 yearly visits, with the first visit as early as possible after birth. The DNBH advocated that each visitor should cover around 300 infants per year in urban areas and 250 infants per year in rural areas (due to longer distances).

Moreover, the DNBH issued very specific guidelines and rules for program content. During their visits, nurses registered the health of each infant on standardized forms. The nurses strongly encouraged mothers to breastfeed and educated them in the three principles of infant care of that time: “calmness, cleanliness, and orderliness.” Moreover, the nurses paid special attention to the correct treatment of additional foods, such as cow’s milk.

The DNBH emphasized the qualifications of the home visitors, insisting that they were

[7] Regional variation in infant care norms—not always favoring breastfeeding—had persisted in the late 19^{th} and early 20^{th} centuries. These regional norms played an important role in explaining both regional variation in infant mortality rates and the high level of mortality rates (Buus, 2001; Løkke, 1998).

[8] Particularly in the choice of control areas, the DNBH was guided by local GPs and avoided low mortality areas.

educated as nurses, had gathered experience in infant care, and obtained additional education by completing training at a new school in Århus. These educational requirements were rarely waived. Moreover, the DNBH issued guidelines for the nurses' cooperation with GPs and midwives. The nurse's responsibility was taking care of healthy infants and referring ill infants to GPs. This referral system left the responsibility for the control of epidemic diseases, for example, to the supervision of doctors. However, this referral system in itself may have constituted a major channel for the effect of the home visiting program, because early contact with all infants allowed the nurses to identify those who were ill and immediately send them to doctors for treatment.

Although highly centralized in its content and procedures, the program was implemented at the municipal level. This decentralized implementation was in the spirit of the time, with the municipalities evolving as the new "big player" in social policies in Denmark, and was central to the liberal and conservative party platforms in the Danish parliament (Buus, 2001). However, the 1937 Act granted a 50 percent refund of costs to municipalities. To be eligible for that refund, the municipalities had to receive accreditation from the DNBH for both their home visiting program and the nurses they wished to hire.

2.3 The Roll-out of the Home Visiting Program

The variation that identifies the effect of the program is the timing of implementation across municipalities. This variation resulted primarily from three factors. First, the DNBH designed educational constraints for new nurses and, second, introduced a time-consuming accreditation process for municipal programs, to better control the inflow of qualified nurses (Buus, 2001).[9] Third, the program remained optional until 1974, so that the preferences of local actors, such as GPs, played a role in the spread of the program.

Materials from the Danish National Archives show that the accreditation process of municipal programs and nurses could last up to one years or more, greatly depending on the quality of the municipal application and likely depending on the DNBH case load. Municipalities had to receive program approval before they could advertise for a visiting nurse. The DNBC reviewed every municipality's plans for the home visiting program and frequently required revisions. While some municipalities chose to employ nurses before receiving official

[9]The inflow of qualified nurses from the Aarhus nursing school was 60 in 1938-40 and 361 in 1941-50 (DNBH, 1970).

DNBH accreditation, they ran the risk of not receiving the full refund.[10]

Figure 3 illustrates the roll-out of the home visiting program across Denmark.[11] The figure shows that from 1937 to 1949 the program expanded gradually. While more towns were among the early movers, considerable variation existed. Copenhagen implemented the program immediately, whereas several surrounding municipalities in the Copenhagen medical district introduced it later. The town of Køge implemented the program much earlier than neighboring Roskilde. Moreover, among the early movers were more remote towns in Jutland, such as Kolding and Sønderborg.

Rural areas also show considerable variation, with early and late movers. Several medical districts applied for district-wide programs that generally covered all rural municipalities. Thus the budgetary capacities of single rural municipalities were unequally important for the introduction of the program across medical districts.[12] Consequently, the considerable variation among neighboring and similar towns and rural areas suggests a high degree of arbitrariness of program roll-out. While my main analysis uses data for towns only, I shall control for the urban-rural dimension in the supplementary analysis, using medical district data.

From the start, although infant survival rates varied greatly across regions, the DNBH did not target the treatment at high mortality regions. Figures 4 and 5 show the year of treatment initiation for towns and medical districts, respectively, as a function of pre-1937 log infant survival (number of infants per 1000 live births that survived the first year of life). Confirming that this dimension was not at the core of the DNBH's strategy, Figure 4 shows that low survival towns were not predominantly found in the early-mover group. While the left panel indicates a weakly positive relationship between high infant survival and early program

[10]Municipalities risked having their nurses not being approved, their guidelines needing revision, or their program being approved as only part-time due to the number of births in the municipality. These factors could lead to a lower refund or no refund. When several municipal programs were assessed as not exclusively focusing on infant care (too few births), the nurses had to be part-time.

[11]The figure is based on the 1970 administrative structure and is thus not fully accurate. In 1970 many small municipalities merged into bigger units, so that the number of municipalities decreased from around 1300 to 277. Figure 3 shows the 1970 municipalities that may contain earlier treated and untreated municipalities. Nevertheless, the figure gives a feel for the geographic spread of the program across Denmark.

[12]Bornholm, Holbæk, and later Frederiksborg, Roskilde, Aarhus and Maribo districts introduced district-wide programs. The dark area in northwest Zealand results from a medical district-wide program for all rural municipalities and a number of towns the medical district of Holbæk. In this district one municipality had been part of the earlier trial, so perhaps its positive experiences with that trial made the Holbæk district move early. However, the same speed does hold not for Vejle and Give, which also took part in the trial. Furthermore, infant mortality rates were much higher in Holbæk's neighboring district Frederiksborg, which did not implement the program until 1947.

implementation, this finding is driven by the major towns (the diamond-shaped symbols in the left panel of Figure 4). These towns were—as argued earlier—most likely to have pre-existing philanthropic interventions such as infant wards. The right panel excludes the major towns and the few small towns with infant wards. As a result, any obvious relationship of pre-treatment mortality and timing of treatment for the remaining towns is eliminated.[13] The same observation holds for Figure 5, which plots subdistricts' entry into treatment and their pre-1937 infant survival rate for all subdistricts as well as only for subdistricts without infant wards.[14]

Overall, personnel constraints (a small and slowly growing pool of qualified nurses), a long and complicated approval process, and differences in the local preferences of key actors (such as GPs and politicians) constitute important explanations for regional differences in the timing of implementation. Thus I assume that the timing of treatment initiation was largely exogenous.

3 Data and Descriptive Statistics

3.1 Data for Short-run Outcomes

To identify the effect of the home visiting program on infant survival, I use various data sources (see also C for a detailed description). First, I have collected unique data on the exact date of implementation of the home visiting program in treatment municipalities for 1937-1949.[15] This information is available in the Danish National Archives from records on the accreditation of municipal programs (The Danish National Archives, various years).[16] I assign treatment status to municipalities according to the following rule: municipalities that implemented the program before July 1 are assigned to be treated in the given year.

[13]The R^2 for the fitted lines in Figure 4, which predicts treatment initiation by using information on baseline infant survival, is 0.01 and 0.00, respectively.

[14]In the subdistrict data I exclude a subdistrict if one of its towns has an infant ward.

[15]As the program changed significantly after the 1940s, I limit my analysis to the years up to 1949. More and more municipalities chose to hire combined infant and school-visiting nurses from 1949 on, the year the school doctor program became compulsory (Bagger, 1960). While the early years witnessed an increase in nurses, the coverage of all municipalities took much longer. For example, as late as in 1952, only 478 of the around 1300 municipalities had visiting nurses. In 1960, 505 municipalities were covered. By the early 1960s, around 1.2 million of Denmark's approximately 4.5 million inhabitants still lived in municipalities without coverage (Bagger, 1960, 1964).

[16]Given that the state refunded some of the expenses to the program, the Ministry of Internal Affairs kept track of the program implementation. Unfortunately, this data does not contain reliable information on the number of nurses at the municipal level.

Municipalities that implemented the program after July 1 are assigned to be treated in the following year.

Second, from the annual DNBH publication *Causes of Death*, I use information on outcomes—most importantly, infant survival rates in the first year of life—at the town level. I then connect this data to the *Danish Statistical Commune Data Archive*, which contains socio-economic control variables at the municipal level (Bentzon, 1975). To proxy for time-varying town characteristics, I compute the yearly percentage of social-democratic voters, the percentage of the population employed in agriculture and industry, the percentage of property and income taxpayers, and the percentage of female population for towns during 1933-1947.

Third, in its annual *Medical Report for the Kingdom of Denmark*, the DNBH published information on a variety of public health issues at the medical subdistrict level, including infant mortality, the number of live births, and the aggregated number of visiting nurses employed in each medical subdistrict's municipalities. Although this data is at a higher level of aggregation than the treatment, it has its merits: First, it also includes rural areas; second, it includes information on specific causes of death.[17] While the causes of death are available only for the general population and not for age groups, for acute enteritis—a condition with the most casualties in the under-one age group—I argue that this measure is a good proxy for infant mortality for the condition. Furthermore, the reports contain information on general characteristics of the medical system of the time at the medical district level: the number of doctors employed in hospitals (1933-1949), of GPs (1940-1949), of midwives (1933-1949), and of hospital beds (1933-1945).

In the analyses that use this subdistrict (and district) data, I assign areas to be treated as soon as the medical report reports a non-zero number of nurses employed. Thus, although only a percentage of the municipalities in the subdistrict is covered, the entire subdistrict is defined as treated.[18] As I pool treated and untreated municipalities, and as not all nurses started their work on January 1, the subdistrict and district data contains more measurement

[17] Although causes of death are also available at the town level, the small size of many towns makes a study of specific causes of infants' death infeasible.

[18] Compliance rates close to 100 percent in treated areas suggest that compliance rates in untreated ones would have been high had they been treated. Using the number of nurses per 1000 live births, I also generate a nurse ratio measure, which reflects treatment intensity. The exogeneity assumption, however, is probably less convincing for the expansion of the program in already treated areas. This expansion could be driven by changes in the infant mortality rate after the program implementation. Thus I prefer the specifications that include a treatment indicator.

error than the town data. Thus I expect my findings from these data to be smaller than the results found in the town sample, and to represent a lower bound of the true effect of the home visiting program on infant mortality.

Tables 2 and 3 show summary statistics for my town sample. While Table 2 compares the five major towns and all remaining towns, Table 3 divides the sample by treatment status. The major towns depart—as previously mentioned—from most Danish towns in important ways. During the period they face lower infant mortality rates, potentially due to pre-existing interventions such as infant wards. Moreover, their inhabitants differ from small town inhabitants in their occupational status, which is likely to coincide with differences in living conditions. The major towns additionally have a higher percentage of income taxpayers, i.e., in the period of an expanding welfare state and increased importance of income tax revenues, they were more likely to have the means to quickly initiate more additional policies.[19]

Table 3 indicates a lower infant mortality rate (and higher survival rate) for treated towns. A similar pattern is visible in Table 4, which is based on subdistrict data. Treated subdistricts have a lower infant mortality rate and consequently a higher survival rate. Moreover, mortality from acute enteritis is almost double the size in untreated areas. The average number of nurses per 1000 live births in treated subdistricts is three. This number conforms to the DNBH guideline of about 300 infants per nurse. Table 5 divides subdistricts in urban and rural ones. It shows that those are similar with respect to infant survival rates and mortality from acute enteritis in the period—despite a higher mean for morbidity from the condition in urban areas. Towns on average have more nurses, while the average number of live births for urban and rural areas are similar, most likely because of a nursing shortage that allowed qualified nurses greater choice of employment (Appendix figure A.2 illustrates this distribution of nurses between urban and rural areas and its development.)

3.2 Data for Long-run Outcomes

While the major focus of this paper is on short-run effects of the home visiting program, I take a first glance at potential long-run returns to universal home visiting. As the home visiting program aimed at improving infant nutrition, and given that a focus on breastfeeding was at the core of the program, a natural starting point for an analysis of long-run benefits

[19] All information on town characteristics (except tax variables) comes from non-yearly data, and I use a linear approximation to arrive at yearly data (see appendix C).

is outcomes that have been shown to correlate with early infant nutrition. I focus on adult height and the probability of being overweight. In accordance with other studies using data for the same period, I define adult overweight as having a Body Mass Index (BMI) of 25 or above (Christensen et al., 1981).

My preliminary analysis of long-run effects is constrained by data availability. Although future analyses should consult population data from the Danish administrative registers when examining other potential long-run benefits for socio-economic and health outcomes, the register data does not contain measures of body weight or height. Thus I turn to data from the Danish military examination, which is obligatory for all able-bodied men aged 18. Unfortunately, for the period I only have access to data on male twins from these examinations for the entire country.[20] Important to note, my identification strategy does not exploit the twin dimension as both twins always are either in the treated or control group.[21]

My sample of male twins eligible for military service contains information on place and date of birth, as well as the year of examination for the birth cohorts 1931-1949. The information on age at examination allows me to distinguish cohort and age effects. The data also contains information on adult height and weight, allowing me to calculate the twins' BMI. I connect the data on outcomes to information on treatment status at the municipal level.

The twin data poses important challenges. During infancy and childhood, twins have most likely faced circumstances that are different from those experienced by singletons, including higher infant mortality. Thus twins that survive until their military examination are to a higher degree selected on underlying health than singleton children. Moreover, although I am not aware of special programs or measures directed at twins during the period, twins likely received more attention from medical professionals. Given their lower birth weight and higher risk factors on average, twins were most likely under special surveillance. Thus twins were subject to treatment by visiting nurses in addition to any other eventual treatments. The question remains whether results based on my twin sample are generalizable. A final concern is the small sample size, which makes it hard to detect any long-run effects of the home visting program.

[20] Military examination records for the region of Zealand are available for all men, and can be used in future analyses for testing the results of this paper. However, as variation at the municipal level is at the core of the identification strategy of this paper, data on only one region will have less usable variation.

[21] I include both male twins—if alive—and all male twins from mixed twin pairs. The percentage of complete twin pairs in my data remains stable over the period.

Appendix Table B.1 shows descriptives for my twin sample for twins in municipalities that implemented the home visiting program from 1937-1949. I estimate all models on this sample and exclude individuals from municipalities that did not implement the program in the period under consideration. Doing so yields in the most comparable control group available. Columns 2 and 3 show summary statistics for the pre- and post-implementation period, respectively. My main sample contains around 4000 individuals. Around half of the men are born after the treatment initiation and around 60 percent are born in urban areas. Comparing pre- and post-implementation individuals shows that mean height increases over the period while mean weight remains stable. This finding partly reflects strong cohort effects for adult height in the period under consideration.[22]

With respect to BMI, similar patterns (as described in earlier studies) emerge: Appendix Table B.1 shows that mean BMI is stable over time, and so is the percentage of men being overweight (BMI>=25). Appendix Table B.2 displays mean BMI, its standard deviation, and year of birth by the year in which the men attended their military examination. Comparing mean BMI and its standard deviation to the figures in a sample of singleton men at their military examination, I find that both remain remarkably stable over time and are very similar across samples (Christensen et al., 1981).[23] Thus my twin sample displays similar summary statistics as a more general sample of recruits of the time. Appendix Table B.3 shows no clear trend in the percentage of overweight by military examination year. This table also shows that until the military examinations in 1956, the BMI information is missing for 30-40 percent of my sample, due to registration practices in the different regions (e.g., on Zealand registration of height and weight started in 1957) (Sørensen and Price, 1990). Thus I test the robustness of my results for overweight by changing the period in my analysis.

4 Empirical Strategy

The decentralised implementation of the home visiting program provides a natural experiment that allows me to evaluate the effect of its implementation. Using variation across towns and medical districts, I estimate the relationship between the log of the infant survival rate and

[22]Appendix Table B.3 illustrates the increase in height for successive birth cohorts.

[23]Table 2 on page 406 in Christensen et al. (1981) shows mean BMI, its standard deviation, and the prevalence of overweight for a sample of singleton men at their military examination between 1943 and 1977. I create my Table B.2 by year of military examination, not birth year, to match this table.

the implementation of the home visiting program from 1933 through 1949:

$$Y_{s,t} = \alpha + \beta \times Treated_{s,t} + \mu_s + \delta_t + \eta_s \times timetrend + \epsilon_{s,t} \tag{1}$$

where $Y_{s,t}$ is the log of infant survival at the town level s for the year t. Unfortunately, the nature of the data does not permit me to analyze heterogeneous effects of the program for different subpopulations, such as first-time or unmarried mothers.

On the right hand side, I include an indicator for each town μ_s and a year fixed effect δ_t. The inclusion of town and year fixed effects allows me to control for all time-invariant town differences and year-specific shocks—such as the effect of the post-war years—that might impact infant survival. The $\epsilon_{s,t}$ is a random error term. The coefficient of interest is β, the coefficient of the treatment indicator taking the value 1 for all periods after implementation of the program and 0 otherwise. This coefficient can roughly be interpreted as the percentage change in infant survival. I cluster all standard errors at the town level. All regressions are weighted by the average number of live births in the towns.

Major threats to identification are parallel interventions and diverging trends. I address these threats in three ways. First, town-specific time trends allow me to capture potential diverging trends across towns in the development of infant survival rates. Second, I restrict my sample of towns to test for the potential confounding impact of pre-existing programs such as infant wards. Third, I add a vector of time-varying town controls $X_{s,t}$ to equation (1). Thus, in my most encompassing specification, I identify the effect of the home visiting program beyond time-invariant town characteristics, year effects, and town-specific time trends and controls. For remaining confounding factors and parallel policies to bias my analyses, they would have to vary in the same geographical and timely pattern as the implementation of the home visiting program.

My regressions for subdistrict and district data are similarly based on equation (1). Here I also include linear time trends, which I allow to differ for urban and rural subdistricts or for each subdistrict. Additionally, I add district-level time-varying controls for the medical system.[24]

[24] To examine the robustness of my results at this level of aggregation, I have also estimated regressions, in which I include a measure for the ratio of nurses per 1000 live births at the subdistrict level. Results are available on request.

For the long-run outcomes height and overweight I estimate, similarly to equation 1,

$$Y_{i,m,t} = \alpha + \beta \times Treated_{m,t} + \gamma \times Age + \mu_m + \delta_t + \eta_{medicaldistrict} \times timetrend + \epsilon_{i,m,t} \quad (2)$$

where $Y_{i,m,t}$ is the outcome of interest for individual i born in municipality m in year t. I transform adult height into birth year-specific z-scores showing the deviation of an individual's height from the mean height in my sample.[25] β is in this case interpreted as changes in unit of standard deviations. I control for age at military examination. β identifies the effect of the home visiting program on treated individuals. I cluster standard errors at the municipality level.

5 Results: The Home Visiting Program and Infant Survival Rates

Figures 6 and 7 provide a first indication of a positive treatment effect of the home visiting program. The figures plot the mean of log infant survival for the years up to and after treatment initiation for all towns and subdistricts, respectively. Both figures shows an upward trend in infant survival for the entire period. Around treatment initiation, a clear jump in the fitted lines—moreso for the town data—is visible.

Table 6 presents the main results for the effect of the home visiting program on log infant survival in year t. While the top panel of Table 6 includes all towns, I restrict my sample by excluding major towns and all towns with infant wards in the lower panels. Controlling for infant ward initiation instead of restricting my sample, I find very similar results for the effect of home visiting.[26] All results are based on town-year observations, standard errors are clustered at the town level, and the regressions are weighted by the average number of live births in each town. All models include town and year fixed effects. Column 2 additionally controls for a linear town-specific time trend.

The results in Table 6 are in line with the graphical evidence. They confirm a positive effect of the introduction of the home visiting program on infant survival rates across specifications. The inclusion of town-specific time trends does not change this conclusion. The choice of

[25] Results are not sensitive to this transformation; results for height in cm are available on request.
[26] Results are available on request.

control group of non-treated towns does not significantly alter the result. I also estimated the regressions excluding towns that never implement the treatment during the period, and find very similar results (columns 3 and 4). I use all towns in the following analyses.

In line with prior expectations, the effect of the home visiting program is stronger for the sample that excludes the major towns. While I find that that the home visiting program increased infant survival by roughly 0.5 percent in my full sample, I find an effect of around 0.8 percent when excluding the major towns. In the most restricted sample of towns in panel three, I find that the program increased infant survival by 0.5-0.6 percent. A likely explanation for this result is pre-existing initiatives, which confound the potential effect of the program in the major towns. These efforts had most likely already decreased mortality from relevant causes and diminished the pool of untreated infants likely to benefit from home visiting. As the major towns on average saw over 4,000 live births annually, the weight assigned to those observations is substantial, explaining their strong influence on the estimate of the effect of the treatment.

Tables 7 and 8 test the robustness of the main result by examining the timing of the program's effect. I estimate a regression in which I include indicators for the years preceding and following treatment initiation for the sample of all towns and for the sample that excludes major towns and minor towns with infant wards. In both tables, I balance my samples, ruling out the possibility that changes in the composition of the sample across years could bias my result.[27] As Table 7 shows, the coefficients for the year dummies are insignificant in the years preceding treatment initiation. In the year of treatment initiation the coefficients turn positive and are significant for the years $t + 0$ through $t + 3$. Very similarly and in accordance with my main result, I find a significant treatment effect only in the years after the introduction of the program for the restricted sample in Table 8. In both tables the effect of the program seems to increase in size over time. This finding may indicate that the program was more effective once firmly established in the local communities.[28]

Table 9 adds to the main results by including a set of town-specific controls for socio-economic background characteristics (1933-1947). Again, I present estimates for the full

[27]The restrictions placed on the sample are motivated by my attempt to balance the trade-off between having a reasonable period around treatment initiation and having a reasonable number of town observations in the balanced sample. Forty-seven towns meet the inclusion criteria in the full town sample.

[28]The absence of a similar pattern in Figure 6 could be related to the figure's being based on an unrestricted sample.

town sample and for samples that exclude the five major towns and infant ward towns.[29] I present specifications that control for the town-year percentage of social-democratic voters and percentage of income taxpayers.[30]

The coefficients for the treatment indicator in Table 9 are extremely similar to the ones in Table 6, i.e., without controls for town characteristics. While the results for the full town sample are insignificant, point estimates for the restricted samples are significant and indistinguishable from the main result. These results, taken together, indicate that the home visiting program was highly effective in the small and medium-sized Danish towns, where the program was very likely the first publicly funded and large-scale intervention for infants and their mothers.

In sum, my results indicate that my main finding of a positive effect of the home visiting program on infant survival rates in Danish towns is driven by majority of (small) Danish towns in most of which the program was the first publicly funded and large scale intervention for infants and their mothers. To complement the analysis based on town data with an analysis for the entire country, and to examine specific causes of death, I turn to more aggregated data on subdistricts and districts. As this data aggregates many municipalities—treated and untreated ones—I expect the effects of the program to appear smaller in this data (as only a percentage of the inhabitants in a subdistrict get treated).

Table 10 presents estimates for the effect of home visiting on the infant survival rate at the subdistrict level. As Table 10 shows, I find similar effects for the program on log infant survival rates when using this aggregated data. The point estimates indicate—in the most flexible specification including subdistrict-specific trends in column 3—that the program increased infant survival by 0.4 percent, i.e., point estimates are similar in size to the full town sample estimates and smaller than the estimates for the restricted town samples.[31]

[29]Estimations that control for infant ward initiation yield in very similar coefficients for the home visiting program indicator. To illustrate the impact of the weighting of the town data, Appendix Table A.2 presents results for the full sample and unweighted data.

[30]As the tax data has yearly information, and together with the data on election results, has the highest quality, I chose these two proxies for town characteristics. Results for other controls are similar and available on request.

[31]The subdistrict data results are very similar for weighted and unweighted data, while at the town level the differences in population size are more extreme, as is the impact of weighting the data. I have also estimated regressions using a ratio measure for the number of nurses per 1000 live births in the subdistrict as my independent variable. I find similar results, i.e., the ratio measure predicts similar increases of infant survival: at the mean of 3 nurses/1000 live births, infant survival increases by around 0.6 percent. However, as the ratio measure is likely to have even more measurement error, it is not surprising that results for the ratio measure are less robust across specifications. Results are available on request.

Table 11 repeats the event study presented for town data in Table 7. To ensure that composition changes in my sample do not drive the results, this table also balances the sample of subdistricts. Table 11 provides further evidence—for the full and the restricted sample—that the program was effective and that its effects on infant survival emerged in year t and $t+1$ and onwards, respectively. The delay observed in the full sample could result from not all nurses registered in the yearly medical reports having started employment on January 1.

Reservations relating to the aggregation of data are necessary for interpreting the results in Table 12, which repeats the analysis at the highest possible level of aggregation—the medical districts—and adds to the full picture by controlling for characteristics of the medical system. Each column in the table refers to the control variables included (number of hospital physicians and midwives, number of GPs and midwives, and number of hospital beds, respectively) and each panel represents the analysis for a different outcome.[32] The periods for the different regressions vary due to data availability.

I analyze the following outcomes: the infant survival rate, the infant mortality rate from acute enteritis, the overall mortality rate (excluding infants), the mortality rate from puerperal fever and lobar pneumonia (pneumonia crouposa), and the stillbirth rate. All rates are per 1000 population, and all outcomes are in logged form. All specifications include district and year fixed effects, and cluster standard errors. Columns 2 and 3 add time trends.[33]

The analysis on data for death rates from specific causes can serve as a falsification test, as the program should not—directly—impact mortality for adults, etc. Although determining which causes of death should and should not be impacted is difficult and data on specific causes is likely to be more noisy, I expect the program to impact infant survival and mortality especially from those infectious disease that can be averted by proper infant nutrition and/or breastfeeding.[34]

[32] I have also regressed the set of district-specific controls on the treatment indicator, year and district fixed effects, and time trends. I find that treatment initiation is uncorrelated to these changes in the number of medical professionals at the time. Results are available on request.

[33] All regressions are weighted by average population in the district, respectively. The one exception is the mortality rate from acute enteritis. Here I assume that the figures reflect predominantly infant deaths. Consequently I calculate the death rate per 1000 live birth and weight the regressions with the number of live births. My results are not sensitive to this procedure.

[34] A parallel falsification test for town data appears in Appendix Table A.3. However, I have fewer available outcome measures at the town level. While the top panel shows a significant effect of the program on the overall survival rate (survival rate for the total population excluding infants), this effect is very small and not robust across specifications.

Looking at cause-specific mortality rates, I find no pattern in the mortality decrease of the period that can be attributed to the program for overall mortality in the over-one age groups, stillbirths, puerperal fever and pneumonia.[35] I find indication for sizeable effects of the home visiting program on mortality from acute enteritis. Given the period's enormous decrease in mortality from acute enteritis, the estimates for the effect of the program that range between 0.24 and 0.75 do not appear unreasonably large. They imply that the program accounted for the largest part of the period's overall decrease of mortality from the condition (the decrease in mortality was around 80 percent in the period). Unfortunately, I cannot examine deaths from acute enteritis with my town data, which would allow me to better pinpoint infant ward towns and their impact on the specific causes of death.

Overall, testing my main result from data for all Danish towns with more aggregated data for the entire country, I find my main conclusions to be robust. As my analysis compares differences in outcomes for areas (towns, subdistricts, or districts, respectively) that implement the program in a given year to outcomes in other areas that do not do so, the treatment and control group should follow a similar trend to make this comparison valid. By controlling for additional time-varying socio-economic controls and features of the medical system, I can rule out the possibility that my estimates pick up the effect of these improvements. Unfortunately, I do not have all control variables at the same level of aggregation.

A number of events deserve special mention, among them other medical advances and public policies of the time. I assume that factors such as improved access to hospitals can be captured by year fixed effects and time trends. Other time-varying improvements that most likely affected all infants' early life conditions include the introduction of new drugs and vaccines (sulfa drugs and other antibiotics developed in the 1930s and 1940s), the introduction of improved access to prenatal care in 1946 (more doctor and midwife visits granted), and the expansion of a school doctor program. None of these policies were implemented parallel to the home visiting program, all of them were implemented at the national level, and only the school doctor program displayed some regional variation in the implementation. Prenatal care and vaccine programs were centrally administered by the DNBH and made available by physicians across the country. The same holds for access to drugs, which was regulated by

[35]Pneumonia was especially dangerous for the very young and the older population. There is strong evidence that the decrease of mortality from this major infant killer was mainly attributable to the "sulfa revolution" (Bhalotra and Venkataramani, 2011).

medical district hospitals and GPs. While urban-rural differences remained in the take-up of vaccines and prenatal care, subdistrict fixed effect and time trends should also account for these differences.

Finally, in principle, when municipalities face budget constraints and constant spending, spending on the home visiting program could take resources away from other policy areas. In such a case, the improvements resulting from the home visiting program would potentially be underestimated, as deterioration in other areas work in the opposite direction. However, factoring in the general background for municipalities and medical districts in the 1930s and especially in the 1940s, I judge this problem to be of minor importance. During the period under consideration, municipal and state spending on social services increased despite difficult economic conditions. The municipal increase in spending was facilitated by an increase in state grants to a number of municipally administered social policies, such as social housing. Furthermore, the introduction of the 1933 financial compensation program across municipalities allowed poorer municipalities to provide new social services (Hansen, 1977; Pedersen, 2009; Jonasen, 1994).

6 The Costs of the Home Visiting Program

This section relates the costs of the home visiting program to the estimated benefits with respect to infant survival rates, thereby examining the cost-effectiveness of the program.[36] This analysis distinguishes itself from a cost-benefit analysis by not assigning a monetary value to the person-years saved by the program. I make a number of simplifying assumptions: I consider only direct costs, i.e., I do not consider indirect costs that could arise from, say, increased child morbidity of surviving infants. Moreover, I consider only estimated immediate benefits in terms of infant mortality decreases (or rather, number of person-years saved), i.e., I do not consider other possible effects, such as fertility responses and long-run benefits.

Appendix C describes in greater detail both the data on costs to the program and the ways in which they have been collected. Appendix Table A.4 illustrates the typical expenditures to visiting nurses in 1945 through three municipalities. Corresponding to the high educational demands on the nurses, most nurses had relatively high seniority when hired as visiting

[36] This section draws heavily on Cutler and Miller (2005).

nurses.[37] Taking these municipal figures as the basis for calculating the expenditure for a "typical" visiting nurse in 1937-1949, I assign the costs of 6900 DKK to each nurse.[38] Appendix Table A.5 shows the aggregated costs per nurse in 1945 for a bigger sample of municipalities. My estimate for the costs per nurse is in the upper part of the distribution and thus a conservative estimate.

Table 13 presents a rough estimate of the cost-effectiveness of the home visiting program. To calculate cost-effectiveness, I use 1941—the mean year of introduction of the nurse program in my medical subdistrict data. Based on my calculations, the costs per nurse in 1941 were 6,002 DKK. In 1941, the national infant survival rate per 1,000 live births was 945, the number of live births was 71,306, and the number of infant deaths was 3,922. Thus my most conservative estimate for the program effect based on town data implies that the program resulted in the prevention of 270 infant deaths. Life expectancy for a newborn from 1941 to 1945 was 65.6 years (for males) and 67.7 (for females) (Statistics Denmark, 2011). Assuming an average life expectancy of 65 years, I calculate the number of person-years saved as around $269.54 \times 65 = 17,520$ in 1941. In 1941, the number of nurses in Denmark was 122, resulting in aggregated costs of 732,347 DKK in 1941. This figure amounts to a cost per life saved of 2,717 DKK in 1941.

For comparability I convert the costs of the program into 2003 and 2010 DKK/USD. I find that the costs of the program per person-year saved was 771/117 (2003) or 882/158 (2010). Cutler and Miller (2005) present similarly generated estimates for the effects of another major public health intervention of the early 20th century that aimed at improving hygienic conditions, namely, clean water technologies in the U.S. Moreover, the analysis from Moehling and Thomasson (2012b) for home visiting in the U.S. constitutes another relevant comparison for my estimates of costs per life saved.

My estimate for the costs per person year saved is—as expected—lower than the estimate from Cutler and Miller (2005), who find that clean water technologies bought an addional person-year at 500 USD (2003 USD). Compared to the estimates in Moehling and Thomasson (2012b)—who calculate that home visiting could save an additional infant at around 20,400

[37]Seniority was defined as years since becoming a general nurse.

[38]I assume the following expenditures for the municipalities per nurse in 1945 DKK: basic wage, 2100; seniority allowance, 1200 (highest seniority); local allowance, 300; civil servant allowance, 1600; yearly-granted allowance, 250; pension contribution, 330; wages for fill-ins, 500; transportation and clothing, 300; telephone, 120; and other, 200.

USD—I find that the Danish program costed only around 10,314 USD per infant death averted (2010 USD). Thus I conclude that the Danish program achieved improvements in infant survival at very modest costs.

A recent review on the development economics literature on the ways in which to fight high disease burdens in developing countries discusses today's challenges and the policies that have been successful in a historical context (Dupas, 2011). Dupas (2011) highlights that, historically, large-scale public health investments have been critical to combat infant mortality and that similar supply side measures might still be important in developing countries. Today a considerable percentage of infant deaths in developing countries is caused by treatable infectious diseases.

Analyzing potential interventions, recent studies from developing countries have shown that early access to health professionals, who can council parents, and identify and treat treatable diseases, can go far in fighting unnecessary deaths. As an example, campaigns for promoting hand washing seem to be successful in preventing high levels of diarrhea, one of the leading causes of death for infants around the world (Wilson and Chandler, 1993; Cairncross et al., 2005). As another example., Gruber et al. (2012) find that improved access to universal health care among the poor in Thailand from the year 2000 onwards had huge effects on infant mortality rates. They also attribute these effects to the better management of treatable diseases.

With high mortality rates and weakly developed health care systems, many developing countries resemble Denmark in the 1930s and 1940s. Despite obvious constraints, the analysis of the costs and the effects of home visiting in Denmark in this period can be informative for policy makers today.

7 A first Glance at Long-Run Health Returns to Home Visiting

Given the negative effect of the home visiting program on infant mortality, the question arises as to whether this short-run benefit translated into longer-run returns for treated cohorts. The negative effects for infant mortality most likely result from two underlying mechanisms related to the infant health distribution and the mortality threshold (Almond, 2006). Being a universal offer and reaching nearly all infants, the program most likely improved overall cohort health in treated municipalities. Thus the program shifted the underlying infant health distribution and should lead to better long-run health outcomes, such as adult height, for treated individuals. However, by focusing on inexpensive ways of improving infant nutrition and hygienic conditions, and on referral of ill infants to GPs, the home visiting program potentially also lowered the mortality threshold in treated municipalities. As a result, more weaker infants from presumably poorer backgrounds survived the first year of life in treated municipalities. This mechanism could result in treated adults being in worse health than their untreated counterparts. Thus health improvements for the entire population of infants have to outweigh the scarring effects (that result from the survival of weaker infants in treatment municipalities) to produce detectable positive long-run effect of home visiting.

This section provides a first glance at potential long-run outcomes that have been shown to correlate with early life nutrition status. Adult height has been used as a proxy for adult health status that is influenced by early life nutrition and disease burden (Bozzoli et al., 2007). Furthermore, several studies have focused on the origins of the current obesity epidemic in Europe and the U.S. and have indicated an important role for early life circumstances, such as infant nutrition (Parsons et al., 1999; Harder et al., 2005; Schack-Nielsen et al., 2010). For Denmark, a number of studies find interesting patterns relating to birth year, i.e., cohort effects: Although mean BMI remains stable in the 1930s and 1940s, the early 1940s see a sharp increase in the share of overweight and obese individuals (Sørensen and Price, 1990; Christensen et al., 1981; Sørensen et al., 1999). This finding strongly suggests the presence of environmental factors to which those cohorts have been exposed and which triggered overweight in individuals with certain genetic dispositions.

While earlier studies have suggested that the implementation of the home visiting program has contributed to the increase in obesity in Denmark (Sørensen et al., 1999), this suggestion has not been tested empirically. The home visiting program's focus on breastfeeding and infant nutrition could play a role explaining increased rates of overweight and obesity—and the exogenous variation in the implementation of the program can help to identify this effect.

Appendix Figure B.1 illustrates the development in mean height and the percentage of overweight men by birth year. The left panel shows a comparison of individuals in treatment municipalities before and after the implementation of the home visiting program. The right panel compares individuals in municipalities that eventually received treatment to individuals in municipalities that did not get treated in the period under consideration. For successive birth cohorts mean height increases similarly over the period in both treatment municipalities and the ones that did not receive treatment. Treated individuals appear to be taller than their untreated cohort peers. For the percentage of individuals with a BMI of 25 or above the two types of municipalities (figure in the right panel) follow a similar trend with the exception of the birth cohorts from the early 1940s. For these cohorts treatment municipalities show a sharp increase of overweight rates which is somewhat delayed in non-treated municipalities. The left panel of Appendix Figure B.1 shows that for almost all birth years treated individuals are more likely to be overweight than untreated ones. However, for both height and overweight differences in pre- and post-treatment means could be due to other differences not accounted for in these graphs.

Appendix Table B.4 presents results for the effect of the home visiting program on adult height (birth year-specific z-scores) and the probability of being overweight at the military examination. Both columns include a municipality of birth and a birth year fixed effect. All standard errors are clustered at the municipality of birth level. Column 2 additionally includes a medical district linear time trend.

The sign of the estimated effect indicates that the home visiting program decreased adult height (by around 0.11 of a standard deviation or 0.7 cm for treated individuals) and increased the probability of being overweight (around 3 percentage points). Moreover, the sign of the point estimates is suggestive for a scarring story, where treated individuals are weak survivors and are actually in worse health than the controls in the long run. However, my estimates for long-run effects are not robust to the inclusion of linear time trends.[39] The size and characteristics of my twin sample pose—as stated above—a major challenge to this analysis of longer-run effects, and future analyses should turn to other data and examine this question further.

[39] Appendix Tables B.5 and B.6 show that the results for overweight are reasonably robust in terms of the size and sign of the effect to limiting the years that I include in the regression and thus not driven by the lack of data for obesity in the earliest birth cohorts.

8 Discussion and Conclusion

A growing literature in economics has turned to the analysis of conditions in infancy to find the origins of health and socio-economic disparities in later life. If early influences have long-lasting consequences, early interventions for improving infant health, such as home visiting, are key policies for combatting adult inequalities.

A considerable percentage of infant deaths in developing countries today is caused by treatable infectious diseases. A recent review of the development economics literature on the ways in which to fight high disease burdens highlights that, historically, large-scale public health investments have been critical for combatting infant mortality, and that similar supply side measures might still be important in developing countries (Dupas, 2011). Home visiting—granting early access to health professionals and better management of treatable diseases—might be one means of preventing unnecessary deaths. With high mortality rates and weakly developed health care systems, many developing countries resemble Denmark in the 1930s and 1940s. Despite obvious constraints, this analysis of the costs and the effects of home visiting in Denmark in this period can be informative for policy makers today.

Using several complementary data sources from Denmark, this paper shows that universal home visiting contributed to the decline in infant mortality rates in the 1930s and 1940s. The main results suggest that the implementation of the home visiting program increased infant first-year survival rates in Danish towns by between 0.5-0.8 percent and was most effective in the majority of small and medium-sized towns. Although the period under consideration features other important changes, the timing and geographic spread of treatment initiation across municipalities facilitates the identification of program effects. My results are robust to a number of tests, among them the inclusion of time trends, and different sets of control variables. Furthermore, and as also stressed in historical research on the causes of infant deaths in given period, my results suggest that one mechanism for the success of the home visiting program was the promotion of proper infant nutrition and breastfeeding, both of which most likely contributed to the period's huge decrease in infant mortality from acute enteritis.

Research on the cost-effectiveness of home visiting interventions suggests that they can be highly cost-effective (Darmstadt et al., 2005). In line with this research, my estimates for Denmark of the 1930s and 1940s show that home visiting increased infant survival rates significantly at modest costs. Similar programs—which are low-technological in their content, improve access to basic medical services, and provide important health information—still have the potential to improve health for many infants around the world.

One advantage of analyzing historical data on home visiting is its potential to shed light on longer-run returns to such programs. Applying the simple model outlined in Almond (2006), I expect potential long-run effects of the home visiting program to be the consequence of two underlying mechanisms: First, improved infant health for the general population of infants, e.g., through shorter sickness spells and increased breastfeeding, could result in improved long-run outcomes for treated individuals. Second, the survival of weak infants, such as infants with low birth weight, could result in scarring effects, i.e., treated individuals could display worse outcomes.

My preliminary analysis of long-run returns to the program carefully suggest that the home visiting program has had an impact over and above its short-run effect on mortality and that scarring mechanisms dominate this effect. While during the period under consideration mean height increased significantly for all birth cohorts, treated male twins seem to be shorter than their untreated counterparts and grow up to have a higher probability of being overweight at about age 19. This finding could suggest that the home visiting program has decreased the mortality threshold in treated municipalities, resulting in the survival of weaker infants. These infants—when compared to their untreated peers—are scarred in later life.

While these long-run results are based on a selected and small twin sample, they are the basis for future analysis of the long-run effects of the program. Earlier studies have hypothesised about the role of the home visiting program in explaining a sharp increase in the prevalence of overweight and obesity rates for birth cohorts in the early 1940s in Denmark. One mechanism for this potential effect of the home visiting program could be the increased probability of survival for low birth weight infants, who earlier have been shown to have a higher probability for adult overweight status. Furthermore, ironically, a potential explanation for the program's contributing to an increase in overweight could be the focus on breastfeeding. At the time, the nurses recommended that mothers breastfeed according to a regular schedule and abstain from feeding during the night. While the nurses' promotion of breastfeeding might have resulted in more women starting to breastfeed, the regularity recommendation might have hindered longer breastfeeding durations. Today's WHO breastfeeding recommendations emphasize more flexible breastfeeding practices. Shorter breastfeeding durations as a result of regularity recommendations might have resulted in the early introduction of complementary food for some infants. A recent study using Danish data suggests this factor as a correlate to overweight in adulthood (Schack-Nielsen et al., 2010). Overall, my results encourage future research on the specific content of the home visiting program and the role of the program in explaining the origins of the obesity epidemic in Denmark.

Finally, future research should use comprehensive Danish administrative register data to examine potential long-run effects of the home visiting program on outcomes such as educational attainment, adult health, and labor market attachment. Research that exploits this data can increase our knowledge of other potential gains of universal home visiting.

References

Almond, D, "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy*, 2006, *114*, 672–712.

_ **and J Currie**, "Chapter 15: Human Capital Development before Age Five," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4, Part 2 of *Handbook of Labor Economics*, Elsevier, 2011, pp. 1315–1486.

_ **and K Chay**, "The Long-Run and Intergenerational Impact of Poor Infant Health: Evidence from Cohorts Born During the Civil Rights Era," *Working paper*, 2006.

Bagger, E, "10 års statistik fra spædbørns- og skolesundhedsplejen - og hvad så? [10 years of statistical material on the home visiting program]," *Tidsskrift for sygeplejersker*, 1960, pp. 207–210.

_ , "Lov om sundhedsplejersker [The Act on the home visiting program]," *Tidsskrift for sygeplejersker*, 1964, pp. 548–555.

Barker, D J, ed., *The fetal and infant origins of adult disease*, British Medical Journal, 1992.

Ben-Shlomo, Y and D Kuh, "A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives," *International Journal of Epidemiology*, 2002, *31* (2), 285–293.

Bentzon, KH, *The Danish Statistical Commune Data Archive - Documentation*, Aarhus University, Insitute of Political Science, 1975.

Bhalotra, S and A Venkataramani, "The captain of the men of death and his shadow: Long-run impacts of early life pneumonia exposure," IZA Working Paper 2011.

Bozzoli, C, AS Deaton, and C Quintana-Domeque, "Child Mortality, Income and Adult Height," Working Paper 12966, National Bureau of Economic Research March 2007.

Buus, H, *Sundhedsplejerskeinstitutionens dannelse: en kulturhistorisk analyse af velfærdsstatens embedsværk [The development of the home visiting program in Denmark.]*, Københavns Universitet, 2001.

Cairncross, S, K Shordt, S Zacharia, and BK Govindan, "What causes sustainable changes in hygiene behaviour? A cross-sectional study from Kerala, India," *Social Science and Medicine*, 2005, *61* (10), 2212 – 2220.

Christensen, U, S Sonne-Holm, and TIA Sørensen, "Constant median body mass index of Danish young men, 1943-1977," *Human Biology*, 1981, *53* (3), 403–10.

Cunha, F and J Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

Cutler, D and G Miller, "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States.," *Demography*, 2005, *42* (1), 1–22.

Darmstadt, Gary L, Zulfiqar A Bhutta, Simon Cousens, Taghreed Adam, Neff Walker, Luc De Bernis, Lancet Neonatal, and Survival Steering, "Evidence-based, cost-effective interventions: how many newborn babies can we save?," *Lance*, 2005, *365* (9463), 977–988.

Delaney, L, M McGovern, and JP Smith, "From Angela's Ashes to the Celtic Tiger: Early Life Conditions and Adult Health in Ireland," IZA Discussion Papers 4548, Institute for the Study of Labor (IZA) November 2009.

DNBH, *Causes of Death in the Kingdom of Denmark*, The Danish National Board of Health, 1933-1950.

_ , *Betænkning Nr 573: Sundhedsplejerske Institutionen [Report of the Commission on the Home Visiting Program]*, The Danish National Board of Health, 1970.

_ , *Medical Report for the Kingdom of Denmark*, The Danish National Board of Health, various years.

Dupas, P, "Health Behavior in Developing Countries," *Annual Review of Economics, Vol. 3, pp. 425-449, 2011*, 2011.

Engle, PL, MM Black, JR Behrman, MC de Mello, PJ Gertler, L Kapiriri, R Martorel, and ME Young, "Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world," *The Lancet*, 2007, *369* (9557), 229–242.

Forsdahl, A, "Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease?," *British Journal of Preventive and Social Medicine*, 1979, *31* (2), 91–95.

Frandsen, KE, *Atlas over Danmarks administrative inddeling efter 1660 [Atlas on the Danish administrative structure after 1660]*, Dansk Historisk Fællesforening, 1984.

Gruber, J, N Hendren, and R Townsend, "Demand and Reimbursement Effects of Healthcare Reform: Health Care Utilization and Infant Mortality in Thailand," Working Paper 17739, National Bureau of Economic Research January 2012.

Hansen, SA, *Økonomisk vækst i Danmark, Bind II: 1914-1975 [Economic growth in Denmark 1914-1974]*, Akademisk Forlag, 1977.

Harder, T, R Bergmann, G Kallischnigg, and A Plagemann, "Duration of breastfeeding and risk of overweight: a meta-analysis.," *Americal Journal of Epidemiology*, 2005, *162* (5), 397–403.

Jonasen, V, *Dansk socialpolitik 1708-1994 [Danish social policy 1708-1994]*, Den sociale Højskole i Aarhus, 1994.

Kamerman, SB and AJ Kahn, "Home Health Visiting in Europe," *The future of children; special edition on Home visiting*, 1993, *3* (3), 39–52.

Løkke, A., *Døden i barndommen-Spædebørnsdødelighed og moderniseringsprocesser i Danmark 1800 til 1920 [Death during childhood: Infant mortality and modernisation in Denmark 1800-1920]*, Gyldendal, 1998.

Løkke, A, "Ro, renlighed og regelmæssighed. Sundhedsoplysning og amning ved Københavns børneplejestationer 1908-1930. [Calmness, cleanliness, and orderliness. Health education and breastfeeding in infant wards in Copenhagen 1908-1930]," *Bibliotek for Læger*, 2008, pp. 474–489.

Loudon, I, *Death in Childbirth. An International Study of Maternal Care and Maternal Mortality 1800-1950*, Clarendon Press Oxford, 1992.

Miller, G, "Women's Suffrage, Political Responsiveness, and Child Survival in American History," *The Quarterly Journal of Economics*, 2008, *123* (3), 1287–1327.

Ministry of the Interior, "Cirkulær til samtlige kommunalbestyrelser angående statens refusion af de af kommunerne I henhold til lov nr 85 af 31.marts 1937 om bekæmpelse af sygelighed og dødelighed blandt børn i det første leveår afholdte udgifter til lønning af sundhedsplejersker," Leaflet 1938.

_ , "RA-IM-1946-53-65 until 53-121," Archive Material 1946.

Moehling, CM and MA Thomasson, "The Political Economy of Saving Mothers and Babies: The Politics of State Participation in the Sheppard-Towner Program," *The Journal of Economic History*, 2012, *72* (01), 75–103.

_ **and** _ , "Saving Babies: The Contribution of Sheppard-Towner to the Decline in Infant Mortality in the 1920s," Working Paper 17996, National Bureau of Economic Research April 2012.

Parsons, TJ, C Power, S Logan, and ICD Summerbel, "Childhood predictors of adult obesity: a systematic review.," *International Journal of obesity related metabolic disorders*, 1999, *23* (Suppl 8), 1–107.

Pedersen, J, *Danmarks økonomiske historie 1910-1960 [Denmark's economic history 1910-1960]*, Multivers, 2009.

Schack-Nielsen, L, TIA Sørensen, EL Mortensen, and KF Michaelsen, "Late introduction of complementary feeding, rather than duration of breastfeeding, may protect against adult overweight.," *American Journal of Clinical Nutrition*, 2010, *91* (3), 619–27.

Sørensen, TIA and RA Price, "Secular trends in body mass index among Danish young men.," *International Journal of Obesity*, 1990, *14* (5), 411–9.

_ **, CT Ekstrøm, and BL Thomsen**, "Development of the obesity epidemic in Denmark: Cohort, time and age effects among boys born 1930-1975," *International Journal of Obesity*, 1999, *23* (7), 693–701.

Statistics Denmark, "Mellemlevetid for en 0-årig, 1940," Statistikbanken 2011.

Strandberg-Larsen, M, MB Nielsen, S Vallgårda, A Krasnik, K Vrangbæk, and E Mossialos, "Denmark: Health system review.," *Health Systems in Transition*, 2007, *9*, 1–164.

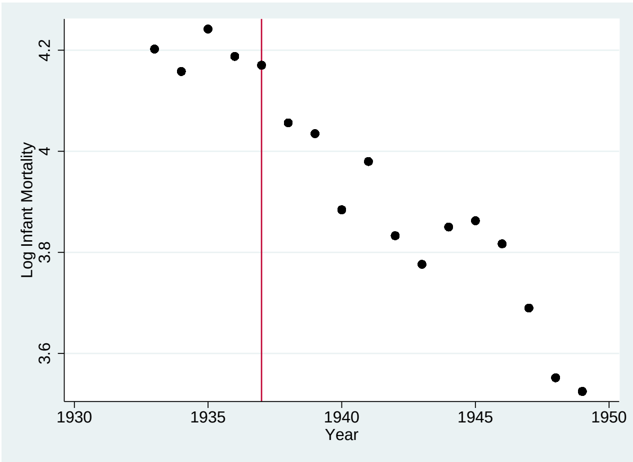
The Danish National Archives, "Cases from the DNBH and the Ministry of Internal Affairs (IM) at the Danish National Archive (RA), 1936-1950, RA-IM-3042, RA-IM-3053 and IM-53," various years.

Vallgårda, S, "Hospitalization of Deliveries: the Change of Place of Birth in Denmark and Sweden from the later Ninteenth Century to 1970," *Medical History*, 1996, *40*, 173–196.

van den Berg, G, M Lindeboom, and F Portrait, "Economic conditions early in life and individual mortality," *American Economic Review*, 2006, *96*, 290–302.

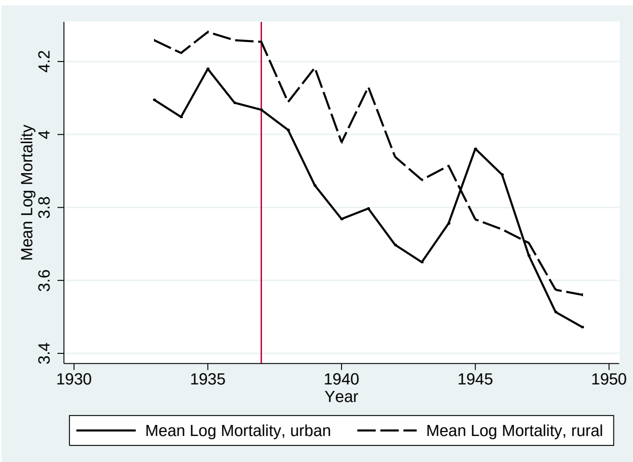
Wilson, JM and GN Chandler, "Sustained improvements in hygiene behaviour amongst village women in Lombok, Indonesia," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 1993, *87* (6), 615 – 616.

Figure 1: Weighted mean of log infant mortality in the medical districts, 1933-1949.



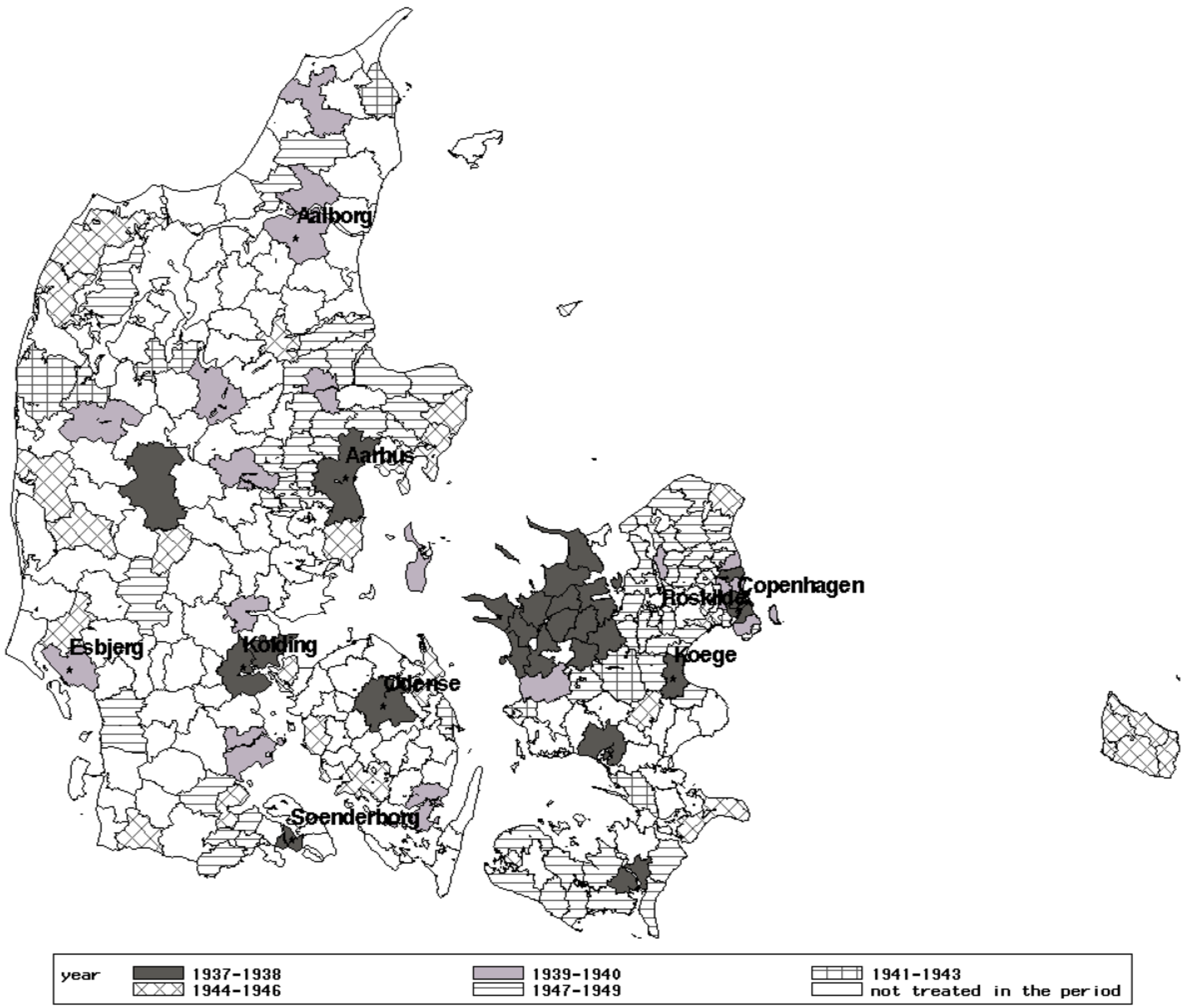
Notes: Mean of log infant mortality rates at the medical district level; weighted by the number of live births in the district and year.

Figure 2: Weighted mean of log infant mortality in urban and rural subdistricts, 1933-1949.



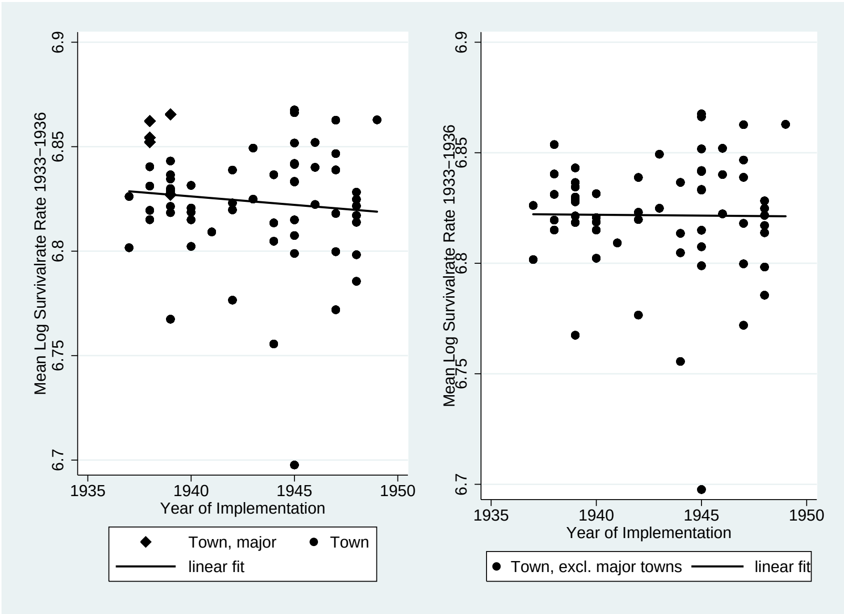
Notes: Mean of log infant mortality rates of urban and rural subdistricts; weighted by number of live births in the subdistrict and year.

Figure 3: Municipalities and their date of entry into treatment, 1937-1949



Notes: Municipality borders based on the 1970s administrative structure, with 278 municipalities.

Figure 4: Year of treatment initiation and log infant survival rate (1933-1936) for all towns that enter treatment and town sample that excludes major towns and infant ward towns, 1937-1949.



Notes: The five major towns are Copenhagen/Frederiksberg, Aarhus, Odense, Aalborg and Esbjerg.

Figure 5: Year of treatment initiation and mean of log infant survival rate (1933-1936) for all subdistricts that enter treatment, and for subdistrict sample that excludes subdistricts with infant wards, 1937-1949.

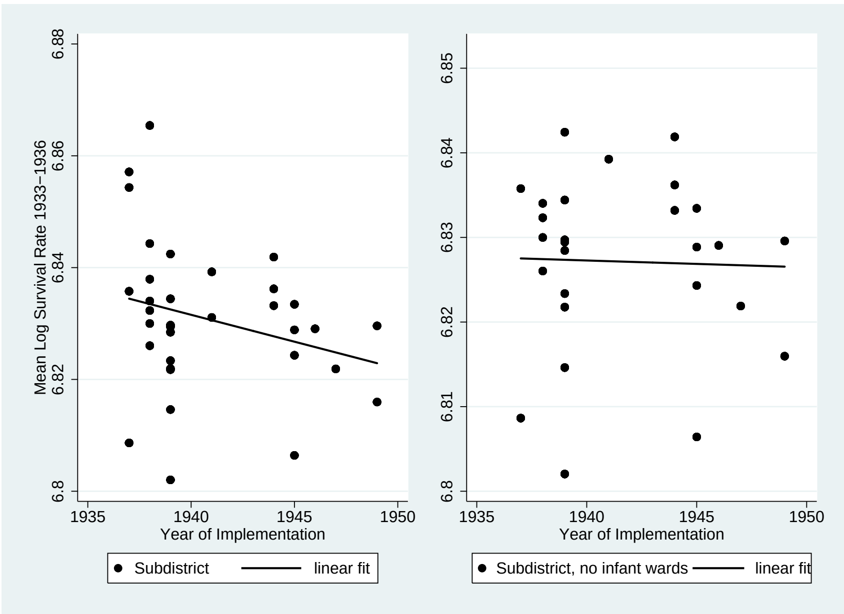
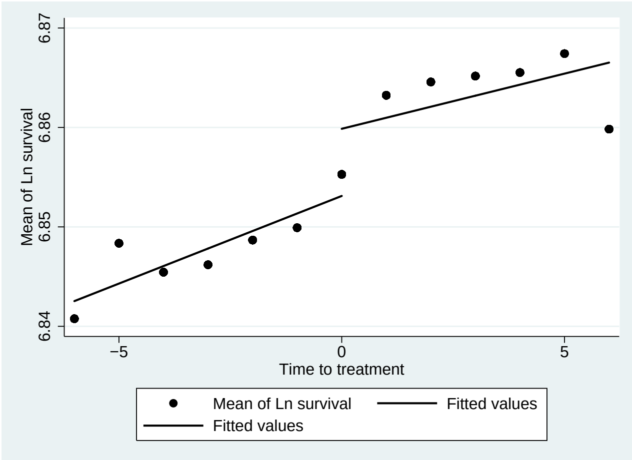
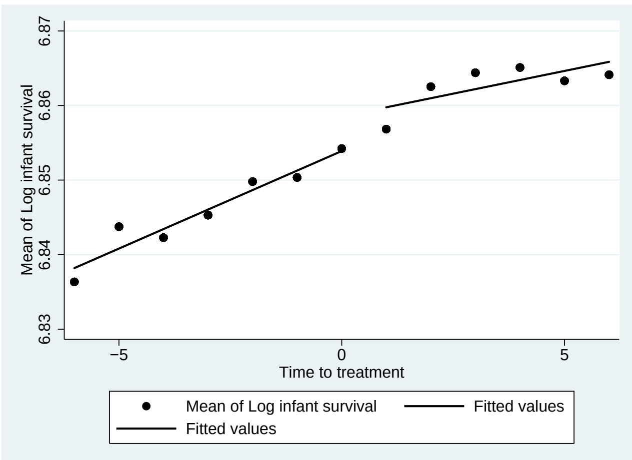


Figure 6: Weighted mean of log infant survival rates in Danish towns by time to treatment.



Notes: Mean of log infant survival rates by time to treatment initiation, weighted by the number of live births in the town and year. No restriction on sample.

Figure 7: Weighted mean of log infant survival rates in subdistricts by time to treatment.



Notes: Mean of log infant survival rates for all subdistricts, weighted by the number of live births in the subdistrict and year. No restriction on sample.

Table 1: Average number of live births and average population in five major metropolitan areas and all other towns, 1933-1949

	Average number of live births	**Average population**
Five major towns	21488.41	1098525
All other towns	14218.53	672659.4

Notes The five major towns are Copenhagen/Frederiksberg, Aarhus, Odense, Aalborg and Esbjerg. Author calculation based on (DNBH, 1933-1950)

Table 2: Summary Statistics for small towns and the five major towns, means and standard deviations, 1933-1949.

	Small towns	*Major towns*
Population, CAD	8304.436	183087.480
	(7369.372)	(237225.937)
No. of town-years	1377	102
Population, imputed	8207.405	181136.271
	(7279.090)	(234608.654)
No. of live births	175.537	3581.402
	(163.098)	(4327.029)
Infant mortality rate	60.692	48.830
	(37.453)	(17.414)
Infant survival rate	939.308	951.170
	(37.453)	(17.414)
No. of stillbirths	6.036	84.657
	(5.928)	(98.174)
Treatment indicator	0.301	0.667
Infant ward	0.040	0.794
Pct. female	0.524	0.528
	(0.018)	(0.020)
Pct. social-democratic voters	0.232	0.246
	(0.058)	(0.045)
Pct. of population in agriculture	0.075	0.023
	(0.057)	(0.027)
Pct. of population in industry	0.391	0.442
	(0.074)	(0.074)
Pct. of income tax payers	0.266	0.332
	(0.070)	(0.066)
Pct. of property tax payers	0.046	0.047
	(0.038)	(0.022)

Notes: See Notes for Table 1; CAD: Causes of Death publication.

Table 3: Summary Statistics for treated and untreated towns, means and standard deviations, 1933-1949

	Treated	*Untreated*
Population, CAD	39111.588	11264.292
	(112641.134)	(47993.590)
No. of town-years	483	996
Population, imputed	42944.554	11708.057
	(118262.834)	(49246.863)
No. of live births	829.306	207.292
	(2182.210)	(781.756)
Infant mortality rate	45.856	66.672
	(20.505)	(40.482)
Infant survival rate	954.144	933.328
	(20.505)	(40.482)
No. of stillbirths	19.977	7.327
	(47.270)	(21.958)
Infant ward	0.178	0.050
Pct. female	0.523	0.524
	(0.013)	(0.020)
Pct. social-democratic voters	0.232	0.234
	(0.047)	(0.061)
Pct. of population in agriculture	0.046	0.081
	(0.035)	(0.060)
Pct. of population in industry	0.430	0.382
	(0.072)	(0.072)
Pct. of income tax payers	0.268	0.272
	(0.084)	(0.067)
Pct. of property tax payers	0.039	0.049
	(0.058)	(0.023)

Notes: CAD: Causes of Death publication.

Table 4: Summary Statistics for treated and untreated subdistricts, means and standard deviations, 1933-1949

	Treated	*Untreated*
Infant mortality rate/1000 live births	46.904	62.822
	(13.962)	(16.372)
No. of subd.-years	311	420
Infant survival rate/1000 live births	953.085	937.194
	(13.952)	(16.320)
Live births	2037.084	1528.602
	(2695.993)	(1170.381)
No. of nurses	6.916	
	(13.143)	
No. of nurses/1000 life births	3.130	
	(1.825)	
Infant ward	0.264	0.088
Sick acute enteritis/1000 live births	792.282	616.132
	(401.874)	(314.928)
Death acute enteritis/1000 live births	3.791	6.269
	(2.932)	(3.526)
Sick puerperal fever/1000 population	0.049	0.066
	(0.050)	(0.055)
Death puerperal fever/1000 population	0.013	0.018
	(0.018)	(0.021)
Deaths/1000 population	10.341	10.342
	(1.448)	(1.465)

Table 5: Summary Statistics for urban and rural subdistricts, means and standard deviations, 1933-1949.

	Urban	*Rural*
Infant mortality rate/1000 live births	56.086	56.012
	(18.551)	(15.870)
No. of subd.-years	374	357
Infant survival rate/1000 live births	943.946	943.963
	(18.480)	(15.872)
Total population	79295.008	98490.353
	(140942.475)	(47385.138)
Live births	1589.243	1908.036
	(2615.118)	(923.916)
Stillbirths	44.321	33.983
	(59.455)	(15.976)
No. of nurses	4.599	1.207
	(11.934)	(4.396)
No. of nurses/1000 life births	2.149	0.475
	(2.086)	(1.349)
Infant ward	0.318	
Sick acute enteritis/1000 live births	791.936	585.409
	(405.311)	(280.951)
Death acute enteritis/1000 live births	5.100	5.335
	(3.821)	(3.143)
Sick puerperal fever/1000 population	0.055	0.066
	(0.060)	(0.045)
Death puerperal fever/1000 population	0.018	0.015
	(0.025)	(0.015)
Deaths/1000 population	11.099	9.548
	(1.358)	(1.087)

Table 6: The effect of the home visiting program on the log infant survival rate at the town level, 1933-1949.

	Full sample		*Treated sample*	
	(1)	(2)	(3)	(4)
		All towns		
Treatment indicator	0.004*	0.005**	0.005*	0.005*
	(0.002)	(0.002)	(0.002)	(0.002)
R^2	0.322	0.453	0.349	0.486
No. of towns	87	87	69	69
		Excluding five major towns		
Treatment indicator	0.008***	0.008***	0.008**	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)
R^2	0.348	0.427	0.384	0.462
No. of towns	81	81	63	63
		Excluding five major towns and infant ward towns		
Treatment indicator	0.006**	0.006*	0.005*	0.005*
	(0.003)	(0.003)	(0.003)	(0.003)
R^2	0.340	0.422	0.379	0.460
No. of towns	77	77	59	59
FE town	yes	yes	yes	yes
FE year	yes	yes	yes	yes
Time trend town level	-	yes	-	yes
Clustered std. errors	yes	yes	yes	yes

Notes: The five major towns are Copenhagen/Frederiksberg, Århus, Odense, Ålborg and Esbjerg. All regressions are weighted by the average number of live births per town. ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table 7: Timing of treatment effects, town data, 1933-1949.

	Log infant survival
Time to treatment: -3	-0.005
	(0.005)
Time to treatment: -2	-0.002
	(0.004)
Time to treatment: 0	0.005*
	(0.003)
Time to treatment: 1	0.012***
	(0.004)
Time to treatment: 2	0.014**
	(0.005)
Time to treatment: 3	0.016***
	(0.006)
R^2	0.482
FE town	yes
FE year	yes
Time trend town level	yes
Clustered std. errors	yes

Notes Sample restricted to towns that contribute at least 5 years both before and after treatment initiation. Regressions include 6 years on both side of year 0, which is the year of treatment initiation. Outcome in the regression is the log infant survival rate in year t. Regression is weighted by the number of live births in town and year. The regression includes dummies for years -3 to +3 around treatment initiation and a dummy for more than 3 years pre and post treatment initiation. The omitted category is year t-1, i.e., the year before treatment initiation. Clustered standard errors in parentheses.

Table 8: Timing of treatment effects, excluding major towns and infant ward towns, 1933-1949.

	Log infant survival
Time to treatment: -3	0.000
	(0.007)
Time to treatment: -2	-0.001
	(0.006)
Time to treatment: 0	0.009**
	(0.004)
Time to treatment: 1	0.013**
	(0.006)
Time to treatment: 2	0.012
	(0.008)
Time to treatment: 3	0.018*
	(0.009)
R^2	0.441
FE town	yes
FE year	yes
Time trend town level	yes
Clustered std. errors	yes

Notes See Notes for Table 1 for town sample details. Sample restricted to towns that contribute at least 5 years both before and after treatment initiation. Regressions include 6 years on both side of year 0, which is the year of treatment initiation. Outcome in the regression is the log infant survival rate in year t. Regression is weighted by the number of live births in town and year. The regression includes dummies for years -3 to +3 around treatment initiation and a dummy for more than 3 years pre and post treatment initiation. The omitted category is year t-1, i.e., the year before treatment initiation. Clustered standard errors in parentheses.

Table 9: Robustness test: The effect of the home visiting program and control variables on the log infant survival rate at the town level, 1933-1947.

	Small towns, no infant wards	
Treatment indicator	0.007**	0.006*
	(0.003)	(0.004)
Pct social-democratic voters	0.052	0.006
	(0.065)	(0.081)
Pct of income tax payers	0.052**	0.058
	(0.020)	(0.041)
R^2	0.323	0.406
No. of towns	68	68
	All towns with non-mission information	
Treatment indicator	0.003	0.003
	(0.003)	(0.003)
Pct social-democratic voters	0.159***	-0.132*
	(0.036)	(0.071)
Pct of income tax payers	0.022	0.082*
	(0.031)	(0.049)
R^2	0.279	0.426
No. of towns	77	77
FE town	yes	yes
FE year	yes	yes
Time trend town level	-	yes
Clustered std. errors	yes	yes

Notes: Tax information is missing for Allinge-Sandvig, Ebeltoft, Fåborg, Hasle, Hobro, Næstved, Silkeborg, Skagen, Skelskoør, Stege. Regressions are weighted by the average number of live births per town. ***significant at the 1 pct level, **significant at the 5 pct level, *significant at the 10 pct level

Table 10: The effect of the home visiting program on the log infant mortality rate and infant survival rate at the subdistrict level, 1933-1949.

	(1)	(2)	(3)
	Log Infant Survival Rate		
Treatment indicator	0.004***	0.007***	0.004***
	(0.002)	(0.001)	(0.001)
R^2	0.644	0.666	0.746
FE subdistrict	yes	yes	yes
FE year	yes	yes	yes
Time trend urban-rural level	-	yes	-
Time trend subdistrict level	-	-	yes
Clustered std. errors	yes	yes	yes

Notes: Number of subdistricts: 43; All regressions are weighted by the subdistrict-year number of live births. ***significant at the 1 pct level, **significant at the 5 pct level, *significant at the 10 pct level

Table 11: Robustness: Timing of treatment effect on log infant survival, subdistrict data, 1933-1949.

	Full sample		No infant wards	
Time to treatment: -3	-0.003	-0.003	-0.006	-0.008
	(0.004)	(0.005)	(0.004)	(0.005)
Time to treatment: -2	0.003	0.003	0.002	0.001
	(0.003)	(0.003)	(0.003)	(0.004)
Time to treatment: 0	0.005	0.006	0.008**	0.008**
	(0.003)	(0.003)	(0.003)	(0.004)
Time to treatment: 1	0.008**	0.008**	0.008*	0.007*
	(0.003)	(0.003)	(0.004)	(0.004)
Time to treatment: 2	0.012**	0.012**	0.012**	0.011*
	(0.005)	(0.005)	(0.005)	(0.005)
Time to treatment: 3	0.014**	0.014**	0.015*	0.014*
	(0.006)	(0.006)	(0.007)	(0.007)
R^2	0.619	0.683	0.653	0.704
FE subdistrict	yes	yes	yes	yes
FE year	yes	yes	yes	yes
Time trend urban-rural level	yes	-	yes	-
Time trend subdistrict level	-	yes	-	yes
Clustered std. errors	yes	yes	yes	yes

Notes Sample restricted to subdistricts that contribute at least 5 years both before and after treatment initiation. Regressions include 6 years on both side of year 0, which is the year of treatment initiation. Outcome in the regressions are the log infant survival rate. Regressions are weighted by the number of live births in subdistrict and year. The regression includes dummies for years -3 to +3 around treatment initiation and a dummy for more than 3 years pre and post treatment initiation. The omitted category is year t-1, i.e., the year before treatment initiation.

Table 12: Robustness: The effect of the home visiting program on various outcomes at the medical district level, additional controls for doctors (in hospitals and GPs), midwives, and number of hospital beds.

	Docs and midwives	*GPs and midwives*	*No. of beds*
		Log infant survival rate	
Treatment indicator	0.001	0.008*	0.000
	(0.002)	(0.004)	(0.002)
R^2	0.769	0.680	0.759
		Log mortality rate, acute enteritis	
Treatment indicator	-0.240**	-0.746***	-0.262**
	(0.102)	(0.203)	(0.116)
R^2	0.472	0.544	0.520
		Log stillbirth rate	
Treatment indicator	-0.009	-0.011	-0.007
	(0.033)	(0.136)	(0.032)
R^2	0.573	0.448	0.539
		Log mortality rate, all	
Treatment indicator	-0.014	-0.022	-0.014
	(0.011)	(0.034)	(0.012)
R^2	0.633	0.803	0.647
		Log mortality rate, puerperal fever	
Treatment indicator	0.269	-0.617**	0.273
	(0.167)	(0.229)	(0.167)
R^2	0.371	0.396	0.366
		Log mortality rate, pneumonia crouposa	
Treatment indicator	-0.084	-0.325	-0.028
	(0.105)	(0.231)	(0.111)
R^2	0.764	0.481	0.760
FE district	yes	yes	yes
FE year	yes	yes	yes

Continued on the next page.

Table 12 *continued.*

	Docs and midwives	*GPs and midwives*	*No. of beds*
District-specific time trend	yes	yes	yes

Notes: "Docs" in first column refers to doctors at hospitals. "No. of beds" in third column refers to hospital beds available. Regression with data at the medical district level; Number of districts=22; Periods for data availability: Midwives: 1933-49; Doctors at hospitals: 1933-46; GPs: 1940-49; Hospital beds: 1933-1945. Clustered standard errors in parentheses. All regressions are weighted by the medical district-year number of live births. ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table 13: Cost-effectiveness of the home visiting program.

	Point estimate	CI low	CI high
Estimated benefits (1941)			
Estimated infant survival increase (percent)	0.4	0.002	0.9
1941 Infant survival increase	269.54	134.77	606.46
1941 Person-years saved	17,519.88	8,759.94	39,419.74
Estimated costs in 1941 DKK			
1941 Expenditures to program	732,347	732,347	732,347
1941 Costs per life saved	2,717.06	5,434.12	1,207.58
Cost-effectiveness of the home visiting program in 2003 DKK/USD			
1941 Costs per life saved DKK	50,082.99	100,165.98	22,259.06
1941 Costs per life saved USD	7,623.43	15,246.86	3,388.19
Costs per person-year saved DKK/USD	771/117	1,541/235	342/52
Cost-effectiveness of the home visiting program in 2010 DKK/USD			
1941 Costs per life saved DKK	57,298.79	114,597.57	25,466.08
1941 Costs per life saved USD	10,313.78	20,627.56	4,583.89
Costs per person-year saved DKK/USD	882/158	1,763/317	392/71

Notes: Statistics Denmark's online calculator at www.dst.dk/prisberegner was used to inflate DKK; 1 DKK in USD (2003): 0.152; 1 DKK in USD (2010): 0.18

A Additional figures and tables

Figure A.1: Danish medical districts in the 1930s

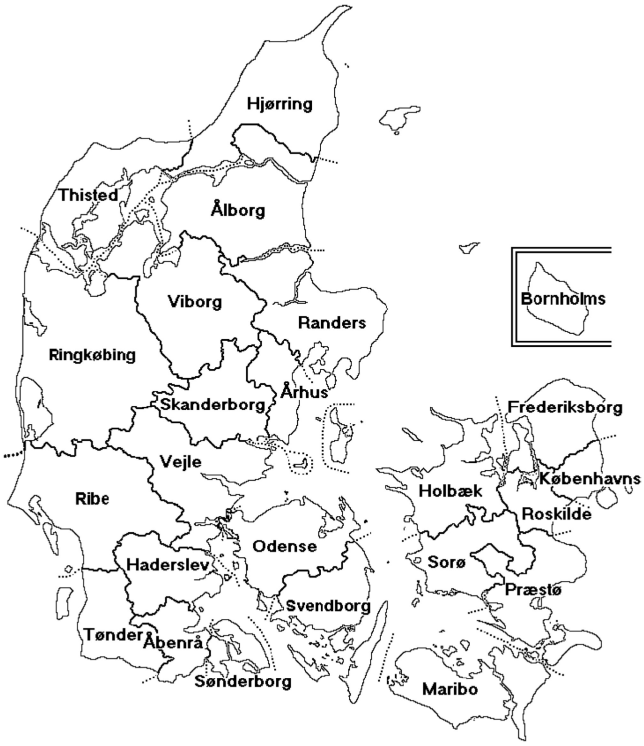


Figure A.2: Number of nurses in urban and rural subdistricts, 1937-1949.

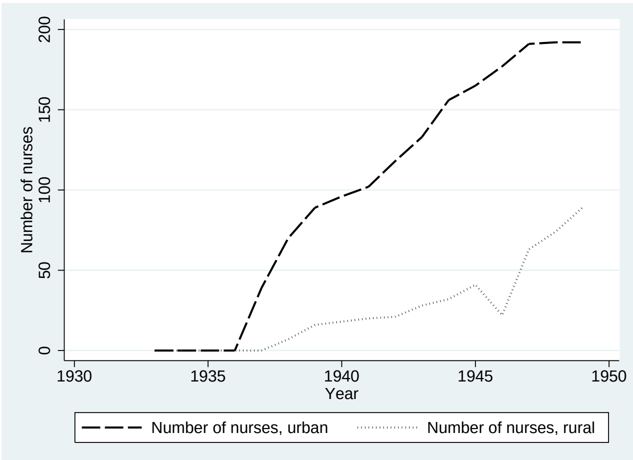


Table A.1: Morbidity and mortality from acute enteritis, DK 1921-1949

Year	Morbidity, infants		Mortality, infants		Mortality, all	
	Number	*Per 100 lb*	*Number*	*Per 1000 lb*	*Number*	*Per 1000 pop.*
1921-1930	5735	8.1	569	8.0	707	0.206
1931-1940	5408	8.2	346	5.2	454	0.123
1941-1945	6925	8.2	282	3.3	401	0.101
1946	7501	7.8	307	3.2	384	0.094
1947	8741	9.5	154	1.7	234	0.056
1948	6677	7.9	100	1.2	153	0.037
1949	6541	8.2	84	1.0	144	0.034

Source: Medical Report for the Kingdom of Denmark 1949. In 1949 Salmonella infections are included. LB: Live births.

Table A.2: Robustness test: The effect of the home visiting program and control variables on the log infant survival rate at the town level, unweighted data, 1933-1947.

	Unweighted, towns with non-missing information	
Treatment indicator	0.010***	0.008**
	(0.003)	(0.004)
Pct social-democratic voters	0.181***	0.180
	(0.067)	(0.116)
Pct of income tax payers	0.025	0.046
	(0.021)	(0.059)
R^2	0.152	0.224
No. of towns	77	77
FE town	yes	yes
FE year	yes	yes
Time trend town level	-	yes

Notes: Tax information is missing for Allinge-Sandvig, Ebeltoft, Faaborg, Hasle, Hobro, Næstved, Silkeborg, Skagen, Skelskør, Stege. Clustered standard errors in parentheses. ***significant at the 1 pct level, **significant at the 5 pct level, *significant at the 10 pct level

Table A.3: Robustness: The effect of the home visiting program on placebo outcomes at the town level, 1933-1947.

Towns with non-missing information		
	Log stillbirth rate	
Treatment indicator	-0.072	-0.059
	(0.047)	(0.042)
R^2	0.279	0.405
	Log survival rate, all	
Treatment indicator	0.000*	0.000**
	(0.000)	(0.000)
R^2	0.166	0.268
No. of towns	77	77
Small town sample, excluding towns with infant wards		
	Log stillbirth rate	
Treatment indicator	-0.014	0.013
	(0.054)	(0.055)
R^2	0.269	0.366
	Log survival rate, all	
Treatment indicator	-0.000	-0.000
	(0.000)	(0.000)
R^2	0.072	0.193
No. of towns	68	68
FE town	yes	yes
FE year	yes	yes
Time trend town level	-	yes

Notes: Additional controls for percentage of income tax payers and percentage of social-democratic voters. Tax information is missing for Allinge-Sandvig, Ebeltoft, Faaborg, Hasle, Hobro, Næstved, Silkeborg, Skagen, Skelskør, Stege. Clustered standard errors in parentheses. Regressions are weighted by the average number of live births per town. ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table A.4: Expenditures for the home visiting program in a sample of municipalities. (1945 Danish crowns)

	Hjørring	Skive	Frederikshavn
No. of nurses (1945)	1	1	1
Year established	1939	1942	1943
Seniority from	1926	1932	1934
Basic wage	2,100	2,100	2,100
Seniority allowance	1,200	1,200	900
Local allowance	318	228	432
Civil servant allowance	1,744.08	1,577.42	1,544.04
Yearly allowance	256.08	209.40	200.04
Muni. pension contributions	396	265.76	0
Wages for “fill-ins”	0	457.62	0
Transportation	100	100	85
Work clothes	0	150	64.59
Telephone	102.15	102.15	165.85
Other	0	269.31	164.66
Sum	6,216.31	6,659.66	5,506.18

Notes: Allowances were centrally designed and dependent on seniority, location, and price adjustments. Pension contributions by municipalities varied between 0 and 10 percent of basic wage and seniority allowance. Other expenditures include expenditures such as printed leaflets for mothers, office rent and cleaning, postal charges, and further education for the nurses. Source: Ministry of the Interior (1946)

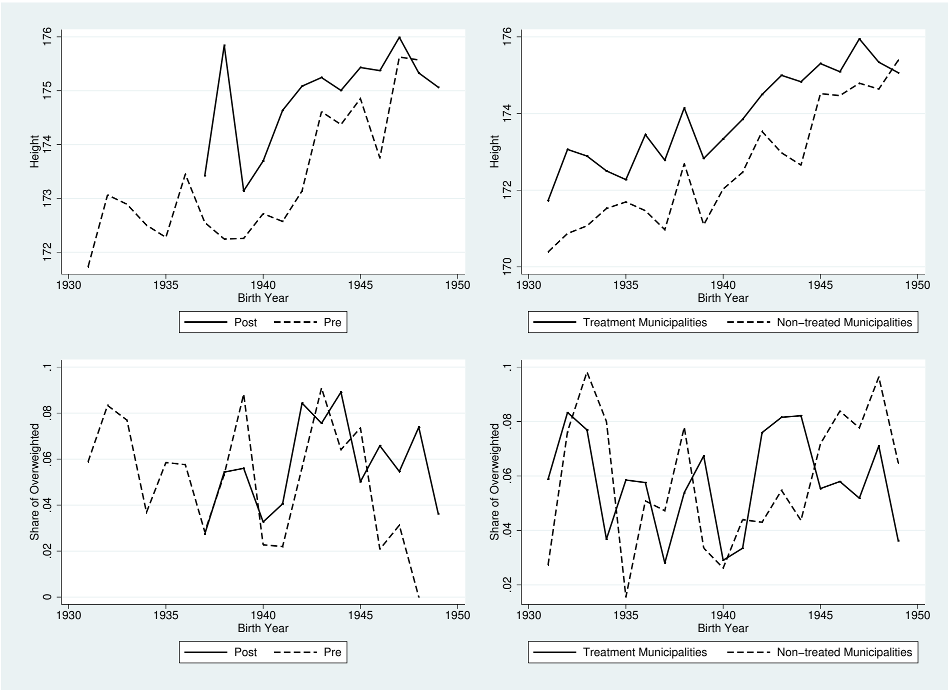
Table A.5: Yearly aggregated expenditures per visiting nurse for a sample of municipalities. (1945 Danish crowns)

Municipality	Costs per nurse
Lyngby-Taarbæk*	6,607.73
Aarhus*	6,905.77
Viborg*	5,866.15
Hjørring	6,216.31
Aalborg*	4,913.90
Gladsaxe*	6,959.58
Holstebro*	6,096.37
Vejlby-Riisskov	6,622.61
Skive	6,659.66
Frederikshavn	5,506.18
Varde	5,469.15
Thisted	7,573.98
Dronningborg**	3,504.27

Notes: Sample of municipalities ordered by date of introduction of the program; *more than one nurse in 1945; **nurse on part-time; Source: Ministry of the Interior (1946)

B Appendix 2: A first glance at long-run outcomes

Figure B.1: Mean height and percentage of overweight individuals for the birth cohorts 1931-1949



Notes: The figures in the left panel only include individuals in treatment municipalities before and after treatment initiation. The figures in the right panel compare individuals in municipalities that implement the treatment in the period considered to individuals in municipalities that do not.

Table B.1: Summary statistics for individuals in treatment municipalities, means and standard deviations, 1933-1949.

	All	*Pre*	*Post*
Height	173.91	172.92	175.01
	6.67	6.57	6.6
N	4563	2397	2166
Z-score for Height	.064	.0376	.096
	1.0	.99	1.01
N	4563	2397	2166
Weight	65.78	65.59	65.96
	8.45	8.43	8.46
N	3942	1788	2154
BMI	21.66	21.86	21.5
	2.32	2.35	2.27
N	3938	1788	2150
Overweight	.058	.055	.06
N	3938	1788	2150
Post	.49	-	1
N	4864	2462	2402
Year of Birth	1940.2	1936.6	1943.9
	5.33	4.3	3.41
N	4864	2462	2402
Year of ME	1959.6	1956.1	1963.2
	5.28	4.22	3.54
N	4864	2462	2402
Age at ME	19.41	19.48	19.33
	1.22	1.16	1.28
N	4864	2462	2402

Table B.2: Mean and Standard Deviation for BMI and Year of Birth by Year of Military Examination (ME)

Year of ME	BMI	Year of birth
1951	21.83	1931.13
	1.62	0.74
1952	22.21	1932.33
	2.96	0.62
1953	21.84	1933.58
	2.22	0.67
1954	22.00	1934.78
	3.43	0.68
1955	21.90	1936.03
	1.94	0.80
1956	21.26	1936.98
	1.78	0.98
1957	21.07	1937.66
	2.14	1.25
1958	21.68	1938.95
	2.64	1.30
1959	21.49	1939.63
	2.63	1.49
1960	21.30	1940.75
	2.14	1.31
1961	21.91	1941.91
	2.18	1.33
1962	21.97	1942.71
	2.25	1.38
1963	21.67	1943.69
	2.03	1.33
1964	21.50	1944.67

Continued on the next page.

Table B.2 *continued.*

Year of ME	BMI	Year of birth
	2.10	1.32
1965	21.74	1945.30
	2.35	1.25
1966	21.56	1946.29
	2.05	1.45
1967	21.55	1947.26
	2.21	1.43
1968	21.70	1948.09
	2.38	1.51
Total	21.67	1940.19
	2.32	5.33

Notes: Individuals in municipalities that implement the treatment.

Table B.3: Average height, percentage of overweight individuals and percentage of individuals with a missing value for BMI by the year of Military Examination

Year of ME	Mean Height	Share of Overweight	Missing BMI
1951	172.16	0.04	0.45
1952	172.92	0.08	0.43
1953	171.84	0.04	0.38
1954	172.42	0.06	0.32
1955	172.70	0.04	0.38
1956	172.79	0.02	0.32
1957	173.05	0.06	0.07
1958	173.14	0.06	0.10
1959	173.25	0.03	0.10
1960	174.27	0.04	0.06
1961	174.86	0.09	0.02
1962	174.34	0.09	0.03
1963	174.48	0.04	0.02
1964	175.44	0.04	0.11
1965	175.33	0.09	0.04
1966	176.17	0.06	0.04
1967	175.75	0.08	0.16
1968	175.58	0.06	0.26
Total	173.90	0.06	0.19

Notes: Individuals in municipalities that implement the treatment in the period.

Table B.4: Long-run effects of the home visiting program: Adult height (birth year-specific z-scores) and Probability of BMI>=25, 1933-1949.

	(1)	(2)
	Adult height (z-score)	
Treatment indicator	-0.108*	-0.086
	(0.062)	(0.067)
	0.085	0.203
R^2	0.146	0.168
N	4546	4015
	Probability of BMI>=25	
Treatment indicator	0.033**	0.029
	(0.016)	(0.018)
	0.034	0.107
R^2	0.124	0.142
N	3933	3524
FE municip.	yes	yes
FE birth year	yes	yes
District-specific time trend	-	yes
Clustered std. errors (municipality)	yes	yes

Notes: Regressions control for age at examination, ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table B.5: Robustness: The effect of the home visiting program on the probability of being overweight (BMI>25)

	(1)	(2)
post1931	0.034**	0.032*
	(0.016)	(0.018)
N	3812	3417
post1932	0.025	0.025
	(0.016)	(0.018)
N	3693	3307
post1933	0.034**	0.035*
	(0.017)	(0.019)
N	3563	3187
post1934	0.035**	0.036*
	(0.018)	(0.020)
N	3400	3044
post1935	0.030	0.032
	(0.019)	(0.022)
N	3211	2875
post1936	0.028	0.028
	(0.021)	(0.027)
N	3020	2695
post1937	0.035	0.030
	(0.025)	(0.033)
N	2805	2506
post1938	0.028	0.014
	(0.030)	(0.035)
N	2640	2366
post1939	0.037	0.019
	(0.031)	(0.038)
N	2447	2192

Continued on the next page.

Table B.5 *continued.*

post1940	0.021	-0.008
	(0.036)	(0.043)
N	2205	1979
FE municipality	yes	yes
FE birth year	yes	yes
District-specific time trend	-	yes
Clustered std. errors	yes	yes

Notes: All coefficients are from regressions for an indicator of overweight status on an indicator for post-implementation. The year for each coefficient indicates the year that I exclude—together with all previous years—from the analysis. Regressions control for age at military examination and only include twins in municipalities that implement the program. ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table B.6: Robustness: The effect of the home visiting program on the probability of being overweight (BMI>25)

	(1)	(2)
pre1938	0.036	0.069
	(0.034)	(0.059)
N	1128	1018
pre1939	0.042	0.072
	(0.028)	(0.049)
N	1293	1158
pre1940	0.038	0.040
	(0.026)	(0.035)
N	1486	1332
pre1941	0.034	0.035
	(0.021)	(0.026)
N	1728	1545
pre1942	0.044**	0.027
	(0.021)	(0.025)
N	1965	1760
pre1943	0.045**	0.024
	(0.020)	(0.026)
N	2202	1958
pre1944	0.036**	0.020
	(0.018)	(0.023)
N	2484	2211
pre1945	0.043**	0.024
	(0.019)	(0.024)
N	2764	2464
pre1946	0.034*	0.020
	(0.017)	(0.021)
N	3068	2739

Continued on the next page.

Table B.6 *continued.*

pre1947	0.035**	0.025
	(0.017)	(0.020)
N	3343	2988
pre1948	0.032**	0.020
	(0.016)	(0.019)
N	3612	3237
pre1949	0.027*	0.015
	(0.015)	(0.018)
N	3795	3404
FE municipality	yes	yes
FE birth year	yes	yes
District-specific time trend	-	yes
Clustered std. errors	yes	yes

Notes: All coefficients are from regressions for an indicator of overweight status on an indicator for post-implementation. The year for each coefficient indicates the year that I exclude—together with all following years—in the analysis. Regressions control for age at military examination and only include twins in municipalities that implement the program. ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

C Overview on data sources

C.1 Municipal level data on outcomes, treatment status, and controls for town characteristics

I use data from the yearly DNBH publication *Causes of Death* from 1933 through 1950 (DNBH, 1933-1950). For the 87 Danish towns of that period, I use data on total population (estimated from Census data), total number of deaths, number of live births, number of infant deaths (0-1 year), and number of stillbirths. I collect data from the *Danish National Archives* on exact timing of treatment initiation in all Danish municipalities for the period 1937 through 1949.

With respect to controls at the town level, the *Danish Statistical Commune Data Archive*, established in the 1960s and 1970s, contains data on local and national elections, and data from the quinquennial Danish Census (Bentzon, 1975). Time periods covered vary for the different data sources. I use electoral data and Census data for the period 1920-1947 for the 87 Danish towns. From the election data I compute the percentage of social-democratic voters for nine elections between 1920 and 1945. I assign lagged percentages from the preceding election to years without election.

From the Census data I use municipal level information on both the total and the female population, the number of those employed in agriculture, and those employed in industry (all available for the years 1930, 1940, 1945). For all variables I use the Census information on the total population to compute percentages. I use a linear approximation to arrive at yearly data for all measures from the Census. To examine the validity of my interpolated measures, I compare my measure for the total municipal population to the data from the *Causes of Death*. As Tables 2 and 3 indicate, my approximated measure is very similar to that in the *Causes of Death* publications. Finally, the data holds yearly information on the number of property tax payers and income tax payers in each municipality for the years 1922 through 1947.

C.2 Medical subdistrict and district data on treatment status, outcomes, and controls for the medical system

The *Medical Reports for the Kingdom of Denmark* for 1933 through 1949 contain information on the year of treatment initiation and the number of nurses for Denmark's medical districts and all urban and rural subdistricts (DNBH, various years). The data also contains the following information at the same level of aggregation: total population, number of live births, number of infant deaths, number of stillbirths, and overall number of deaths. Moreover, the yearly publications contain information on specific diseases (morbidity and mortality), among them pneumonia, influenza, severe enteritis, and puerperal fever. Unfortunately, the mortality data for specific causes is not available by age group at the district level. Given that for acute enteritis, overall deaths are predominantly driven by infant deaths, I assume that all deaths are in the 0-1 age group. I calculate mortality rates by accounting for the size of the total population in each district. Finally, the medical district data also contains information on characteristics of the medical system.

C.3 Data on costs to the program

As, to my best knowledge, there exists no published data for either municipalities or medical districts on spending to the home visiting program for the years considered here, I have collected expenditure data in the National Archives for a sample of municipalities (Ministry of the Interior, 1946). Given that municipalities were entitled to refunds only for expenditures described by the Ministry, and designed their programs accordingly, an analysis of a sample of municipalities offers a solid overview on direct program costs.

As the materials at the National Archives were not collected for the purpose pursued here, I focus on materials from 1945 and pick a sample of municipalities with available and well-structured refund documents.[40] I chose 1945 because by that year most municipalities were using a uniform refund sheet that had been introduced around 1943. In 1945, the payment system for the visiting nurses was structured as described in a leaflet from the Ministry from 1938 (Ministry of the Interior, 1938):

[40] The materials are ordered by year and municipality only, i.e. all topics/correspondence of the Ministry with the single municipalities (e.g., on refunds, new nurses, pension questions, guidelines for nurses' work) are filed together and not all materials appear to have been conserved.

- A basic wage at 2,100 to 3,300 DKK per year, dependent on a seniority allowance of 300 DKK granted every third year,
- Several allowances for working clothes, transportation, and yearly price-adjusted allowances according to the regulations for civil servants,
- Optional municipal pension contributions between 5 and 10 percent of the basic wage and the seniority allowance;
- Compensation for nurses' expenditures for telephone, offices, other materials.

The Effect of Caesarean Section for Babies in Breech Presentation on Child and Mother Health. Evidence from a Regression Discontinuity Design

Vibeke Myrup Jensen and Miriam Wüst[*]

Abstract

Since the 1970s, the use of Caesarean section (CS) for childbirth has increased considerably. The costs and benefits of this development are subject to debate. While CS potentially increases costs for the health care system, improved (or worsened) health outcomes for CS babies and their mothers have to be considered and are of economic relevance in the short and long run. Given that mothers who have a CS are a select group, we use a fuzzy regression discontinuity (RD) design to evaluate the health effects of CS use. In 2000, the Lancet published the influential "Term Breech Trial" (TBT) study. The TBT concluded that planned CS is superior to planned vaginal delivery. We exploit a sharp and discontinuous increase in the Danish CS percentage for babies in breech presentation around the dissemination of the TBT results to identify the effect of CS use on the marginal breech baby. While we find indication for CS use decreasing the probability of having a low APGAR score at five minutes for the marginal child, in general we find no strong health effects on the outcomes that we study. These outcomes include child hospitalizations and outpatient contacts, as well as mothers' post-birth complications. Given these findings we conclude that for the marginal children and mothers (who are a select and relatively healthy group), CS use has no negative short-run health effects either.

[*]The authors acknowledge financial support by the Danish Agency for Science, Technology and Innovation through a grant to the Graduate School for Integration, Production and Welfare (Wüst) and grant number 09-065167 (Jensen). We thank Henrik Nyholm, Nabanita Datta Gupta, Tor Eriksson, Paul Bingley, Mette Ejrnæs, Søren Leth-Petersen, Rafael Lalive, and seminar participants at the Department of Economics and Business, Aarhus University and SFI for helpful comments. Some work on this paper was done during Miriam Wüst's stay at Columbia University, NYC, and benefited greatly from discussions with fellow graduate students Maya Rossin-Slater and Katherine Meckel.

1 Introduction

Since the 1970s, many developed countries experienced an increase in the use of Caesarean section (CS) for child birth. In Denmark, for example, the overall CS rate increased from around 13 percent in the early 1980s to 20.4 percent in 2004 (The Danish National Board of Health, 2005), and in the U.S. the overall CS rate was even higher at 29.1 percent in 2004 (Declercq et al., 2006). While some medical conditions of the mother or the baby offer a clear-cut indication for the use of a CS, critics have argued that this huge increase in CS use indicates the transfer of the procedure to patients for whom the medical indication is less distinct (Shearer, 1993).[1]

The economics literature has mainly focused on direct costs for the medical system induced by increased CS use, costs that can be substantial. Gruber and Owings (1996) refer to estimates from the Health Insurance Association of America stating that, in the U.S., a CS costs 66 percent more than a vaginal delivery. Nevertheless, estimates of the direct costs of CS use that narrowly focus on the reimbursement of the medical procedure must be considered with care. As remuneration systems vary across countries, the costs of a CS may vary as well. Moreover, when estimating the additional costs of a CS, the counterfactual to compare the costs of the procedure to is often not well defined. A complicated natural birth might not be much cheaper than a CS. Thus, the monetary benefits of performing more vaginal births might be overestimated.[2]

As highlighted in the existing economics literature, to analyze the costs and benefits of CS, researchers should also consider other factors, most importantly patients' health. A common finding in the medical literature is that maternal mortality and morbidity rates are higher after CS deliveries and that mothers face higher health risks in following pregnancies, among other things because of an increased risk of a consecutive CS. As to date, little is known about the *causal* short- and long-run effects of CS on patients' health. One major problem complicating the analysis is selection, i.e., mothers who have a CS are a select group. Factors such as, e.g., the mother's and baby's health condition, parity and the size of the baby, and physicians' practice are important for determining the mode of delivery. Given the selection into CS

[1]Main diagnostic groups that often or always lead to elective CS use include multiple births, breech presentation, placenta previa, earlier CS, high risk of emergency CS due to pregnancy complications (The Danish National Board of Health, 2005a).

[2]Gruber and Owings (1996) acknowledge that their estimate of the increased costs due to CS use is an upper bound as the counterfactual birth most likely not is an uncomplicated one.

according to among other things underlying health, we lack a credible counterfactual for CS mothers. Thus the effects of CS cannot be identified by a simple comparison of outcomes for CS mothers to mothers who deliver naturally. By exploiting a unique natural experiment that altered CS rates for breech babies in Denmark, this paper estimates the effect of performed CS on this distinct group of mothers and their babies.

Full breech position at term means that the baby has not turned head down in the womb by week 37 of the pregnancy. While breech position is related to premature birth (because babies move around in the womb more actively before term), among babies at term breech position is present in 3-4 percent of all births (The Danish National Board of Health, 2005a; Tharin et al., 2011). Those breech babies constitute an as good as random subgroup of all births, i.e., it is unclear why some babies do not turn head-down in the last part of the pregnancy.[3] Breech pregnancies have contributed significantly to the rise in CS rates in recent decades, i.e., for the group of breech babies CS rates have increased considerably (Shearer, 1993). In 2005, breech babies accounted for 4.4 percent of all births in Denmark and breech babies born by CS constituted around 20 percent of all performed CS in Denmark (own calculation based on data from The Danish National Board of Health, 2005).

While breech position is present in a constant percentage of all babies (and there is no reason to believe that this will change), breech births are not readily comparable to births with the baby in cephalic presentation (i.e., head-down position). Given the positioning of the baby, vaginal breech births are on average more complicated than births with the baby in cephalic presentation. As a result, on average breech babies at term have an elevated risk of elective CS (to avoid birth complications) and emergency CS during labor.

The elevated risk of poorer birth outcomes for breech babies at term (due to complicated vaginal birth, oxygen deficiency during labor, or injuries due to extraction of the baby) has long been an area of concern for obstetricians (OBs). The continuing debate about the adequate mode of delivery for breech babies led in the mid 1990s to the initiation of the "Term Breech Trial" (TBT). To decide on the adequate mode of delivery for breech babies, the TBT—a large multi-centre and multi-country study—randomly allocated mothers with

[3]Similarly, we do not know why most babies turn around. While breech has not been shown to be determined by one single factor, rare conditions that correlate with breech at term are congenital anomalies,placenta praevia, tumors and the amount of amniotic fluid. However, in most breech cases none of these conditions is present and the baby has not turned for unknown reasons. These cases are the ones where doubt about the optimal mode of delivery remains.

babies in breech position at term to either planned vaginal birth or planned CS (Hannah et al., 2000). In October 2000, The Lancet published the results of this multi-centre multi-country trial. The trial concluded that planned CS is superior to planned vaginal birth with respect to child neonatal mortality and morbidity. However, as concerns about the TBT study design and analysis remain, as longer-run follow-ups of the TBT show no significant differences between groups, and as a number of country-specific observational studies have at most shown minimal differences in short-run outcomes for breech babies according to the mode of delivery, the TBT conclusions remain challenged and a number of critics have questioned the recommendations of the TBT (e.g., Kotaska, 2004; Glezerman, 2006).

Nevertheless, the TBT publication had an enormous impact on CS rates for breech babies in several countries, among them Denmark. In Denmark, the probability of CS for the overall populations of breech babies increased from around 80 to 95 percent in response to the dissemination of the TBT results (own calculations based on the National Patient Registry). Tharin et al. (2011) use Danish administrative register data on all breech births and illustrate that the CS percentage for breech babies increased significantly after the TBT. Furthermore, they find that health outcomes for babies after the TBT and in the planned CS group are more favorable.

Both the TBT and the Danish study by Tharin et al. (2011) perform an intention-to-treat analysis (ITT) that compares outcomes by *planned* mode of delivery. Tharin et al. (2011) compare mean outcomes for breech babies born before and after the TBT publication and in the two periods according to the planned mode of delivery. Comparing risk ratios, they find—in accordance with the TBT—better birth outcomes in the period after the TBT. Comparing babies by planned mode of delivery, the study additionally finds a lower perinatal mortality in the post-publication period for children born by elective CS, as well as fewer low APGAR score babies and fewer neonatal intensive care hospitalizations for this group.

Given considerable non-compliance in the TBT and imperfect compliance in the register study by Tharin and co-authors, another important treatment effect—the average treatment effect on the treated (TOT)—is not identified in this analysis. This present paper contributes to the debate on the health effects of CS by focusing on this (local) treatment on the treated effect. We exploit the discontinuous increase in CS probability around the TBT dissemination to estimate the health effects of actual CS use for the marginal CS breech baby, i.e., the breech baby who would have been born naturally before the dissemination of the TBT results. This marginal group is of great importance and policy relevance in the light of debates on increasing CS rates in comparably healthy populations and on weak medical indication.

Denmark is ideally suited for our study for three reasons. First, to identify the local treatment effect of CS on treated breech babies, we use a fuzzy regression discontinuity design (RD). The fuzzy RD design is equivalent to using the child's date of birth as an instrument for CS use (Angrist and Pischke, 2008; van der Klaauw, 2008; Lee and Lemieux, 2010). In our analysis we use high-quality administrative data which allows us to identify breech babies, their exact date of birth (i.e., their assignment to treatment) and relevant outcomes with minimal error. Second, while countries like the U.S. had very high CS rates for breech babies already before the TBT, the Danish CS rate for breech babies was at a high level but had the potential to increase. We focus our analysis on babies with higher parity than one because those babies are the ones that were impacted by the change of practice (as shown in section 4). Third, given that CS was already a standard and safe procedure before the publication of TBT and several procedures were in place for discovered breech babies (most importantly attempted external version), we are not worried about technological changes coinciding with the TBT publication. By examining a number of possible threats to identification at the cut-off, we provide evidence for this statement in section 4.

Our analysis confirms a strong first stage, i.e., giving birth after the cut-off—the dissemination of the TBT results among Danish physicians on December 4, 2000—predicts CS probability for breech mothers well across specifications. We expect positive effects of CS use for child health as the TBT resulted in an increased selection into CS, i.e., the marginal CS mother and baby are relatively healthy. In accordance to this expectation, our analysis confirms a positive health effect of CS use for breech babies at birth: we find that breech babies born by CS have a lower probability of having a low APGAR score at five minutes (i.e., an APGAR score of seven or below). Furthermore, we find some indication of a decrease in the child's probability of having an above average number of general practitioner (GP) visits in the first two years of life. As GPs are the primary access to the health care system in Denmark, fewer GP visits could indicate persistent positive health effects of CS use for the marginal breech baby.

Although the graphical analysis of our data also indicates a decrease in the probability of experiencing serious perinatal child morbidity, we fail to identify statistically significant effects for this outcome. Moreover, we find no significant effects of CS use for other child health outcomes early in life—measured as hospitalizations, outpatient visits to hospitals, and specialised physicians visits—all of which are very imprecisely estimated. Additionally, especially results for health care utilization measures that include longer periods of the child's life are very small in size. These small point estimates indicate that there are no persistent effects for these outcomes. Finally, for short-run mother health outcomes—the likelihood of having an infection after birth and an indicator for a set of rare maternal post-birth complications—we find no significant results and point estimates are also very small. While this finding suggests no increased short-run health risks for CS mothers, we lack precision in our estimates for these rare post-birth complications.

In sum, the stricter selection of mothers for vaginal birth and increased use of CS after the TBT dissemination has improved an immediate health outcomes, namely the five minutes APGAR score, and there is indication for its decreasing the likelihood of having an above average number of GP visits early in life for the marginal breech baby. However, we find no effects for the marginal breech baby in our analysis for other health care utilization measures. These measures are related to more serious health episodes that require contact with hospitals. The APGAR score has been shown to correlate with certain later life health outcomes such as child morbidity and cognition. Therefore, our future work will focus on examining how much of this initial health effect can be traced in long-run returns in these outcomes (measured as diagnoses and not as health care utilization).

The paper proceeds as follows. Section 2 presents background literature and discusses several potential mechanisms behind increased CS use and medical procedure use more generally. Moreover, section 2 presents the necessary background on the medical system in Denmark and the TBT. Section 3 presents the empirical methods and section 4 presents the data and a detailed graphical analysis. Section 5 contains our estimation results, section 6 discusses our results and concludes.

2 Literature and Background

This section presents the related literature on the increase of CS use and its determinants. Moreover, it presents the relevant background information on the Danish medical system and the Danish guidelines for treatment of breech babies. Finally, it gives necessary information about the TBT and outlines how the TBT changed the practice of Danish physicians with respect to breech babies.

2.1 Related Literature

Since the 1970s, the increase in CS use has been subject to scientific and political debate. Importantly, while considerable doubt remains about the beneficial effects of the intervention with respect to child and mother outcomes, physicians have embraced the wide use of the technology for childbirth (Shearer, 1993). As a number of studies emphasize, changes of the underlying birth-giving population of mothers cannot solely account for the huge increase in procedure use (Declercq et al., 2006). This point is underlined by the observation that a considerable percentage of the overall increase of CS use is driven by increases inside specific diagnosis groups—such as breech (Gregory et al., 1998; Shearer, 1993).

Important explanations for increasing CS rates are technological innovations resulting in improvements of the procedure itself and other technologies such as monitoring of the child's heart rate (continuous cardiotocography (CTG)). Several randomised trials have shown that CTG use increases CS rates while it does not lead to better outcomes such as reduced prevalence of cerebral palsy, infant mortality or improved other standard measures of neonatal well-being (Zarko et al., 2006).[4]

In addition to technological innovations, several studies have highlighted a number of other factors that contribute to the increase of CS rates. Accounting for maternal risk profiles, Baicker et al. (2006) find considerable remaining geographic variation in CS rates and their increase across U.S. counties. They label this unexplained variation "physician style". In a similar vein, Epstein and Nicholson (2005) illustrate that physician style—their beliefs and preferences—matters for explaining variation in CS rates after accounting for other observable characteristics of the physician and the area of practice.

Another set of economic studies has focused on the potential role of physician-induced demand, which predicts that physicians exploit their agency and perform procedures because of expected financial gains (Gruber and Owings, 1996; Gruber et al., 1999; Grant, 2009; Triunfo and Rossi, 2009).[5] Gruber and Owings (1996) and Gruber et al. (1999) find that

[4]CTG has though been shown to decrease the probability of neonatal seizures (ibid.).

[5]Physicians preferences for leisure might be another non-monetary factor that explains physician-induced

reimbursement programs and rules explain physicians' use of CS, a procedure that—in the U.S.—is reimbursed with a higher amount than a natural birth. Using the same U.S. data, Grant (2009) reaches a similar conclusion although he finds smaller effects than Gruber and co-authors.

Finally, and also studied in a U.S. context, increased CS use could be driven by physicians' exposure to liability rules. Currie and MacLeod (2008) discuss the notion of "defensive medicine"—where physicians aim at reducing the risk of legal liability—and the impact of this behavior on childbirth practices in the U.S. They find that certain types of tort reforms (reform of the liability of physicians) increase the use of procedures—among them CS—while others decrease procedure use.

Related to the literature on CS use, a number of studies have focused on the role of newly available information in the context of medical procedure use in general—with a greater focus on the patient's role than the role of physicians. These studies suggest that new information matters for procedure use and open up the question whether this is also the case for CS use.

Price and Simon (2009) find that new information from medical journals on the negative effects of the attempt of a vaginal birth after CS (VBAC) decreased VBAC rates significantly in the U.S. Anderberg et al. (2011) show the impact of (questionable and later withdrawn) evidence on the danger of the measles, mumps, and rubella (MMR) vaccine on parental response in the UK. They find that in response to a heated debate in the media, parents decreased their support for the vaccine and vaccination rates decreased. Both papers focus on patients' response to widely available information—the media took up both topics and discussed them extensively—and illustrate that well-educated patients seem to react more rapidly to newly available information.

Using a RD design, Del Bono, Francesconi and Best (2011) examine the impact of new medical information on potential health risks of the "third generation" pill on women's fertility decisions. Focusing on behavioral changes on the side of the women, they find that both conception rates, abortion rates and birth rates increased as a consequence of the pill scare case. Interestingly, in contrast to the findings of Price and Simon (2009) and Anderberg et al. (2011), in their application the reaction of young mothers and mothers with low socio-economic class is strongest.

demand in contexts, where physicians do not have direct financial incentives, e.g., the Danish context. See Rochut (2010) for an analysis on French and Swiss data.

Our study contributes to these strands of literature on CS use and the impact of new information in two ways. First, a natural extension to the research on increased procedure use is research on the health consequences of this increase. The health consequences of CS are rarely considered in the economics literature. One exception is Currie and MacLeod (2008), who find that increased use of CS after tort reform does not coincide with improved infant health. Given that most studies on the health effects of CS are observational studies and given that we are only aware of one large randomised controlled trial (RCT)—the TBT—that examined the health effects of (planned) CS, we contribute to the literature by focusing on the causal effect of actually performed CS on health outcomes.[6] While we arguably examine the health effects of CS use on the distinct group of breech babies at term, we have credible identification for our local treatment effects on the treated. Breech babies are a small but distinct and persistently present group of births and thus our results have clear policy implications. Furthermore, our paper illustrates the use of the RD design to examine important questions on medical procedure use and its consequences.

Second, we examine the use of CS in a context where financial incentives for physicians are at most modest and indirect and thus we highlight the importance of the impact of newly available information for medical procedure use. While previous studies have focused on patients' response to information, our study examines a case where the newly released information was mainly subject to an expert debate. Thus our study sheds light on the question as to what extent new information influences the supply side, i.e. physicians' practices.

2.2 Background: Relevant Features of the Danish Medical System

This section gives an overview over important features of the Danish medical system and the guidelines for treatment of breech babies. In Denmark, the National Board of Health (DNBH) issues national recommendations for the content of prenatal care. Universal prenatal care is provided free of charge. Mothers are referred to hospitals by their GPs in the first trimester of the pregnancy. From there on, the pregnancy is monitored by GPs, midwives and—if necessary—by OBs at the chosen hospital. While mothers in principle are free to chose their hospital, close to all mothers are referred to the closest hospital according to mothers' residence. As breech presentation is first diagnosed late in the pregnancy, mothers cannot

[6] In Hofmeyr et al. (2003) two earlier RCTs are included in the systematic review. However, these two studies are very small and the TBT accounts for most of the observations by far.

chose their hospital for prenatal care according to their knowledge on different hospitals' policies regarding breech births.[7] Furthermore, close to all mothers give birth at the hospital chosen early in their pregnancy (see the graphical evidence for mothers with breech babies in section 4).

In Denmark, OBs are not paid according to the procedures they perform but are employed at hospitals and receive a wage. Thus, in principle, physicians have no financial incentives to perform a CS.[8] Today Danish hospitals, however, are increasingly reimbursed according to the DRG-system (Diagnosis Related Groups). This activity-based system implies that hospitals potentially have incentives to perform different procedures and/or "upcode" their activities. Upcoding means that physicians assign higher reimbursed diagnoses to patients to increase revenues. In our study, however, this performance-based reimbursement system should not play an important role for two main reasons: first, in the case of vaginal birth vs CS for breech babies the likelihood of upcoding or adjusting CS rates to increase revenues is minimal as the reimbursement for a "complicated" vaginal birth is very similar to the reimbursement for a CS. Second, while the percentage of the hospitals' revenues that are allocated according to the DRG-system has increased over time, in 1999 it was just 10 percent and only increased to 20 percent by 2004, i.e. after the period we consider in this paper.[9]

Already before the TBT publication in 2000, breech presentation was a topic of great attention. The Danish Society of Obstetricians and Gynaecologists (DSOG) provided detailed guidelines for the handling of detected breech pregnancies (see, e.g., Danish Society of Obstetrics and Gynaecology, 1998). Breech pregnancies have a good chance of being detected by trained midwives in the last part of the third trimester. As ultrasound diagnostics are not routinely performed after week 20 of the pregnancy, breech babies are usually detected when midwives examine the mothers' wombs externally and listen to the babies' heart rates. If the midwife suspects breech position, the mother is referred to ultrasound diagnostics. Nevertheless, around 15-20 percent of breech babies remain undetected until the onset of labor (many

[7]While the differences in hospitals' handling of and CS rate for breech, as well as their reaction to the TBT, could be an additional source of variation, we have too few breech births per hospital to pursue an analysis based on hospital variation. While there might have been differences before the TBT in hospitals' CS rates for breech, all hospitals reacted with an increase of their CS rates.

[8]Physicians could face incentives with respect to leisure time as CS are easier to schedule.

[9]Moreover, the major part of the activity-based reimbursement system was used for hospital activity in specific patient groups (such as specific surgical patients and patients from other counties than the hospital's county) (Ministry of the Interior, 2003).

of them have most likely turned around in the womb often and thus they are hard to detect).

According to the DSOG guidelines, the Danish practice for detected breech babies in the 1990s was the following: physicians and midwives monitored the pregnancy closely and physicians attempted to perform an external cephalic version, i.e., an attempt to turn the baby before the onset of labor. In general, around half of the attempts to turn the baby around were (and still are) successful. For breech pregnancies with an unsuccessful external cephalic version, medical professionals and mothers decided on the mode of delivery, i.e., whether the birth should be an attempted vaginal birth or an elective CS.

In order to be eligible for attempted vaginal birth, the DSOG determined the following set of criteria in their 1998 guidelines: adequate pelvic diameter, estimated birth weight of the baby below 4000g, frank or complete breech position,[10] and exhaustive information of the mother on risks and benefits of the procedure. While in Denmark the default for births with cephalic position is that only midwives are present during labor and that OBs only take part in the case of complications, for all vaginal birth attempts for breech babies the DSOG guidelines required the presence of an experienced OB during labor. Furthermore, the hospital had to have access to paediatricians and a specialized intensive care unit, and the medical staff had to comply closely to clearly defined procedures for the handling of labor (including guidelines for amongst other things the inducement of labor, the use of analgesics, the maximum duration of labor, and the suitable manoeuvres for delivering the baby).

2.3 The Term Breech Trial and its Impact on CS Rates in Denmark

The TBT was a large multi-country multi-center RCT that was motivated by considerable disagreement about the adequate method of delivery for most breech babies at term (Hannah et al., 2000). This disagreement was partly due to the lack of evidence from RCTs. On October 21, 2000 The Lancet published the results of the TBT (Hannah et al., 2000). The fast-tracked article was published only six months after the termination of randomization of the last women included in the trial.[11]

[10]Frank breech means baby's hips are flexed and knees are extended. Complete breech means that baby's hips and knees are flexed but feet are not below the baby's buttocks.(Hannah et al., 2000)

[11]The Trial was ended prematurely at an interim analysis that showed significant differences between treatment and control group. The Lancet fast-tracked the publication of the 2000 article featuring the first TBT results. This handling of the TBT results indicates the importance the Lancet attributed to them. However, as the fast publication made it impossible to account for longer-run effects and settle important questions (such as the interpretation of results found for countries with high or low perinatal mortality), the fast-tracking of

The TBT included 2083 women from 121 centres in 26 high- and low-perinatal mortality countries during 1997-2000.[12] The TBT protocol included singleton babies in frank or complete breech position at term, below 4000g, without fetal abnormalities, and without other indications for a CS such as placenta praevia. At each center, women who met the inclusion criteria were randomly allocated to either an attempted vaginal birth or a planned CS.[13] The trial investigated the effects of CS on infant and mother health on an ITT basis. The compliance rates in the TBT were 90.4 percent for planned CS and 56.4 percent for vaginal delivery. The researchers of the TBT did not conduct an analysis of the treatment effect on the treated but exclusively focused on the ITT analysis. Thus, our study—that focuses on the effect of CS for the marginal group of treated mothers—supplements the ITT findings.

The authors of the TBT concluded that planned CS is superior to planned vaginal birth for babies in breech position who meet the inclusion criteria of the TBT. They found significant lower risks of perinatal mortality, neonatal mortality and serious neonatal morbidity for babies in the planned CS group. Additionally, the authors found bigger effects of planned CS for child health in the low-perinatal mortality countries. Furthermore, the study found no difference between the attempted vaginal birth vs the planned CS group in terms of maternal morbidity or mortality in the short run. A systematic review from 2003—including TBT data—concluded slightly differently, i.e., found a small elevated risk of maternal morbidity for mothers with planned CS (Hofmeyr et al., 2003).

While the TBTs quest was to “determine whether planned CS was better than planned vaginal birth fore selected fetuses in the breech presentation at term” (Hannah et al., 2000, p.1375), today, a number of critics of the TBT challenge the conclusions of the trial or at least their external validity (see among others Turner, 2006; Glezerman, 2006; Kotaska, 2004). They point out that—opposed to the detailed protocol of the TBT—not all mothers had an obstetric physician present during labor, twins ended up being included in trial, and different countries had different practice with respect to factors such as attempts of external cephalic version before labor.

an article with the importance of the TBT has been criticised in the following (Bewley and Shennan, 2006).

[12]The countries with low perinatal mortality rates were: Australia, Canada, Chile, Denmark, Finland, Germany, Israel, Netherlands, New Zealand, Poland, Portugal, Romania, Switzerland, UK, USA, Yugoslavia. The countries with high perinatal mortality rates were: Argentina, Brazil, Egypt, India, Jordan, Mexico, Pakistan, Palestine, South Africa and Zimbabwe.

[13]Interestingly, some women must have been in labor at the time of randomization and informed consent. In this case "the CS was undertaken as soon as possible” (Hannah et al., 2000, p.1376).

With respect to research methods, the importance of the "serious morbidity" measure of the TBT has been challenged as it contained a number of different outcomes with rather different consequences for longer-run health of the baby. The TBT did not consider differences among centres and between the medical systems of countries, i.e., there was no stratification according to the two levels in the randomization process. Later reassessment of death cases in the TBT furthermore cast doubt on the relation of these cases to the method of delivery. The short follow-up time of the trial and power issues restricted the possibility to identify longer-run effects. Given the huge difference in standard of care across countries, the external validity of the trial for a country like Denmark with a high quality of medical care and strict criteria for attempted vaginal birth has been questioned (Goffinet et al., 2006). Finally, the considerable percentage of non-compliance among trial participants has not been accounted for in the ITT analysis of the trial data. This factor is even more important as in countries like Denmark only few women agreed to be randomized in the first place.[14]

Nevertheless, the TBT was a ground-breaking trial and its up to date over 500 registered cites in the Web of Science give an impression of the importance of the study.[15] At the time of publication, the TBT had a major impact on national guidelines for the handling of breech pregnancies and on the practice of physicians (Turner, 2006). For example, in the Netherlands the probability of CS for breech babies increased from 50 to 80 percent around the TBT publication (Rietberg et al., 2005). In Australia and New Zealand, 72 percent of OBs reported before the publication of the TBT that they routinely offered vaginal births for uncomplicated singleton breech pregnancies. After the TBT, only 20 percent of OBs reported doing so (Phipps et al., 2003). Also for France an increase in CS rates for breech babies at term as a consequence of the TBT has been documented (Carayol et al., 2007).

For Denmark, several sources strongly suggest that the TBT had major impact on the CS rate for breech babies. First, only few weeks after the TBT publication the DSOG scheduled an extraordinary meeting on December 4, 2000. At this meeting around 200 OBs, gynaecologists and midwives discussed the TBT publication of October 21, 2000 (Danish Society of Obstetrics and Gynaecology, 2001). The December meeting disseminated the findings of the TBT to all physicians and hospitals in Denmark (Clausen, 2003) and constitutes the event that changed the "best practice" for breech pregnancies in Denmark.

[14]Denmark contributed with one woman who agreed to randomization to the TBT.
[15]Number of cites as to November 2011.

Second, the DSOG discussed the TBT on several of their annual meetings (The Sandbjerg meetings in the years 2001, 2003, and 2011). As the DSOG issues national guidelines for physicians' practices, these annual meetings and their expert working groups are essential for Danish physicians. The 2003 Sandbjerg report summarizes the survey response from Danish maternity wards in 1999 and 2001 on their practices for breech pregnancies. Twenty-eight wards (78 percent) answered the survey in both years. Of those, twenty-two wards (79 percent) reported that the TBT publication affected their policy for breech pregnancies. Six out of 28 wards recommended elective CS per default for breech positions before the TBT whereas 18 out of 28 wards recommended CS after the TBT publication.[16]

Third, although no change in national DSOG guidelines occurred in the year 2000/2001, guidelines changed at the hospital level. As an example, the 2001 guidelines from the fifth largest maternity ward in Denmark show, that the results of the TBT were included and resulted in women now having to meet eight inclusion criteria for an attempted vaginal delivery (Aalborg Hospital, 2001). The Aalborg guidelines state that "we [the maternity ward] cannot continue to present attempted vaginal birth as a safe alternative to elective CS. We have to put the numbers on the table. Most women will chose elective CS in this situation [i.e., if the TBT results are presented]" (Aalborg Hospital, 2001, p. 4; authors' explanations in brackets). Additionally, further anecdotal evidence collected among medical professionals also confirms major behavioral changes of physicians and midwives as a result of the TBT and its dissemination on the DSOG December meeting (Clausen, 2003). Fourth, also the analysis in Tharin et al. (2011) mentioned earlier confirms a strong first stage, i.e. an increase in CS percentages around the TBT dissemination.

Taken together, all available information suggests that the dissemination of the TBT results to all OBs and midwives in Denmark changed the best practice for breech births and did so rapidly. Given the institutional setting, this change of practice was not driven by new financial incentives but rather than that by newly available RCT evidence from a highly ranked journal. Given that the topic was heavily debated among experts, the rapid change in CS probability was driven by changed physicians' practice rather than mothers' request.[17]

[16]No published data or published report available. According to MD Henrik Nyholm—one of the key members of the TBT discussions at the Sandbjerg meetings—the TBT was a driving force behind this change towards more elective CS for breech babies (Nyholm, personal interview 2011).

[17]This statement is supported by the difficulties that we experienced when searching for information on the TBT and its impact in Denmark. We have not found evidence for broad media coverage of the TBT around

3 Empirical Methods[18]

To identify the effect of CS use, this paper exploits the discontinuity in the probability of experiencing a CS for Danish mothers with breech pregnancies around the dissemination of the TBT results. Thus we use a fuzzy regression discontinuity design (RD design). In the following we outline the RD design as used in our application.

In the Potential Outcome Framework we can for each individual i define the two outcomes $Y_i(1)$ if exposed (i.e., having a CS) and $Y_i(0)$ if unexposed (i.e., not having a CS) to treatment. As we do not observe any individual in both states—treated and untreated—we cannot directly estimate the treatment effect for mother i, i.e., we do not observe the difference of the two potential outcomes $Y_i(1) - Y_i(0)$.

To overcome this essential evaluation problem, a convenient but problematic approach is to compare the outcomes of exposed and unexposed mothers using regression models. This comparison of average outcomes that are observed for mothers with and without CS results in an estimate of the average treatment effect given by

$$E[Y_i(1)|treated] - E[Y_i(0)|untreated] \tag{1}$$

If mothers who have a CS and mothers who do not have a CS differ in baseline characteristics or if mothers sort into treatment according to expected gains, the estimate of the treatment effect based on equation (1) will be biased. While the RD design does not offer randomization to eliminate these two potential sources of bias, it provides what could be called a local randomisation (Lee and Lemieux, 2010).

The RD design exploits a discontinuity in treatment status that is generated by a cut-off in some observed assignment variable X. In our application, breech births before December 4, 2000, are in the control group, while breech births after this date are in the treatment group. Thus the RD design identifies the average treatment effect locally at the cut-off as

$$E[Y_i(1) - Y_i(0)|X_i = c] \tag{2}$$

For this comparison to be valid, we rely on the *local continuity assumption* which states

its dissemination to medical professionals in Denmark.

[18]This section draws heavily on Lee and Lemieux (2010).

that individuals just below and above the cut-off have similar potential outcomes in the absence of treatment. Furthermore, other characteristics than treatment status develop smoothly through the cut-off, i.e. we assume that treated and untreated individuals close to the cut-off only differ in their value of the forcing variable X and are otherwise comparable. In our application the continuity assumption implies that around the cut-off breech mothers of given characteristics (e.g., health status, educational level or age) would have had a similar probability of experiencing a CS in absence of the TBT and their outcomes would be similar as well. Furthermore, mothers' characteristics other than CS probability are smooth through the cut-off c. Therefore, close to the cut-off c, we can in principle use a simple comparison of the outcomes of individuals on either side to identify the average treatment effect (van der Klaauw, 2008; Lee and Lemieux, 2010). Close to the cut-off c of our forcing variable X,

$$\lim_{x \downarrow c} E[Y_i(1)|X_i = x] - \lim_{x \uparrow c} E[Y_i(0)|X_i = x] \tag{3}$$

gives us the average treatment effect in a sharp RD design.

While "sharp" means that all individuals on one side of the cut-off are controls while all individuals on the other side are treated, our analysis uses a "fuzzy" RD design, i.e., the change in treatment probability at the cut-off is smaller than one. While the dissemination of the TBT results changed mothers' CS probability, the probability increased by a number smaller than one (breech mothers who gave birth before the cut-off already had a risk of CS.)

While a sharp RD design identifies the effect of the treatment directly by estimating the size of the jump in the outcomes Y around the discontinuity c, a fuzzy design is comparable to an instrumental variable approach (Angrist and Pischke, 2008; Imbens and Lemieux, 2007). Giving birth before or just after the publication assigns mothers to different regimes with respect to physicians' threshold of performing a CS. Moreover, as the risk of CS varies across women according to their characteristics, i.e., as women are selected for CS according to their expected gains, the local conditional independence assumption is violated. In this situation we estimate a local average treatment effect (LATE) for the group of complying mothers that changes CS status because of the dissemination of the TBT results. We therefore estimate

$$E[Y_i(1) - Y_i(0)|complier, X_i = c] \tag{4}$$

for women at the cut-off.

Although the focus on the locality of the RD design is crucial for identification, when it comes to the estimation of treatment effects, we have to extrapolate away from the cut-off. Strictly speaking, at the cut-off c there is no data to estimate a treatment effect on. However, the farer away from the cut-off we move, the less credible it is that we meet the criteria for a valid RD analysis. As a consequence, we compare different estimation strategies and carefully document the ways in which we cut the estimation sample.

In our main analysis estimate the effect of CS using two-stage least squares. Our second stage equation is:[19]

$$Y_i = \alpha_2 + \beta_2 \times \hat{CS} + f(forcing) + f(forcing, \hat{CS}) + \delta_2 \times Z_i + \epsilon_{2i} \tag{5}$$

where $\hat{CS}$ is the predicted probability for CS, the $forcing$ variable is calender time in days, Z_i are mother and child-specific controls, and ϵ_{2i} is a random error term. We center the forcing variable around 0. Furthermore, to allow for different slopes on both sides of the cut-off, we also interact the forcing variable with the treatment CS (Angrist and Pischke, 2008). The key assumption is that $f()$ is a smooth and continuous function throughout the cut-off. We discuss the choice of $f()$ below.

Our first stage equation for CS_i—estimating the jump in treatment probability at the cut-off—looks accordingly

$$CS_i = \alpha_1 + \beta_1 \times c + g(forcing) + g(forcing, c) + \delta_1 \times Z_i + \epsilon_{1i} \tag{6}$$

Given that we have to estimate our regressions on data in a larger neighborhood of c, the choice of the function $f()$ (and $g()$) is crucial for a valid RD analysis. We present results for different specifications motivated by a graphical analysis of the data. Our main analysis relies on specifications that include second order polynomials. To assess the sensitivity of results to the specification of $f()$ we show alternatives to this choice. We also present estimates for different windows of data around the cut-off to examine how close to the cut-off we can restrict our data window and still have a strong first stage. We expect point estimates to remain similar though less precise for smaller windows.[20] Additionally, we compare our parametric

[19]We estimate both our first and second stage using the same order polynomials.

[20]As we will see, we face potential weak instrument problems for very tight windows around the cut-off.

results to results obtained by local linear regressions using a rectangular kernel as suggested among others in Lee and Lemieux (2010).[21] To examine the importance of bandwidth choice, we present estimates using the rule of thumb bandwidth (RoT) suggested in Fan and Gijbels (1996) and double this bandwidth.[22]

To examine the validity of our RD design, we present an extensive graphical analysis of our data: we confirm the strength of our first stage both by showing that the jump in treatment probability for breech mothers is not sensitive to a reasonable variation of bin width and by performing a set of local linear regressions with different bandwidths of data. To rule out that mothers manipulate their assignment to treatment or control group—which we find highly unlikely to begin with—we graph the density of our forcing variable around the cut-off. We also inspect graphs for mothers' background characteristics, placebo reforms and groups, and other technologies to rule out jumps around the cut-off. These jumps if present could indicate that the discontinuity design is not valid.

4 Data and Graphical Analysis

4.1 Data

We use administrative register data on the universe of singleton births from 1997 through 2006. We combine several of the Danish administrative registers to obtain information on maternal background and health, child mode of delivery and outcomes.

To identify all live-born breech babies in the population data, we use diagnoses given to mothers throughout pregnancy and at birth, and recorded in the National Patient Register.[23] We exclude babies that have a breech diagnosis but have successfully been turned around before birth.[24] We further exclude breech babies who are born premature, i.e., born before

Given the very flexible specifications and the small window, this is not surprising.

[21]Kernel choice does not alter our main results.

[22]We estimate the RoT bandwidth separately for each side of the cut-off and follow the procedure described in Fan and Gijbels (1996) and Lee and Lemieux (2009) for a rectangular kernel, i.e. we use a quartic regression function to calculate

$$hRoT = 2.702 \times \left(\frac{\tilde{\sigma}^2 \times R}{\sum_{i=1}^{N} \{\tilde{m''}(x_i)\}^2} \right)^{\frac{1}{5}} \tag{7}$$

, where $\tilde{m''}()$ is the second derivative of this quartic regression of Y on the forcing variable, $\tilde{\sigma}$ is the regression's standard error and R is the data range.

[23]Breech babies are defined by the ICD10 codes DO321, DO641, DO641A

[24]A successful external cephalic version is registered with the operation code KMAB10

completed 36 weeks of gestation. This exclusion reflects the fact that physicians–before and after the TBT—usually do not attempt a vaginal birth for premature babies. We also exclude births from wards with less than 100 births in the period as these very small wards are not specialized maternity wards.[25] Our results are not sensitive to this exclusion.[26] Finally, as Appendix Table A.1 illustrates, due to the strict selection criteria for vaginal breech birth already before the TBT, first-time mothers with breech do not experience a discontinuous change in CS probability. Thus we restrict our final estimation sample to births with higher parity than one. At this point we only include live births but we will examine stillbirths in future work. We refer to the births described here as our sample of breech babies in the following sections.

The register data contains detailed information on mother's health and characteristics, method and hospital of delivery, and mother and child outcomes. Furthermore, the data contains the child's date of birth. As the RD design induces quasi-randomization around the cut-off, control variables for mother and child characteristics should not change the conclusions of our analysis. However, they can improve the precision of our estimates and are important with respect to examining the validity of the RD analysis (Lee and Lemieux, 2010). We use information on maternal education and age at childbirth, pregnancy-related health (complications/diagnoses other than breech during pregnancy), and an indicator for child sex.

For the method of delivery, the register data contains information on whether the mother had a CS and whether the CS was an elective or an acute one. Acute CS are defined as a CS performed less than eight hours after it was scheduled. However, in the period considered in this paper the coding for the two different types of CS changes and a considerable percentage of CS is not unambiguously defined as either elective or acute. Furthermore, while the TBT compared children's outcomes by planned method of delivery, we conduct our main analysis for the actual method of delivery (CS vs completed vaginal birth).[27] We find this interesting in its own right as the TBT most likely changed physicians' behavior for both types of CS:

[25]This restriction results in the omission of 100 births.

[26]We do not account for babies' congenital diseases and anomalies. We will examine the robustness of our results to their exclusion in future analyses. However, for the RD analysis—as long as continuous in the neighborhood of c—the inclusion of babies with anomalies should not matter much as they most likely were delivered by a CS both before and after the TBT.

[27]The data reveals that both types of CS increased for breech babies after the TBT. However, the percentage of acute CS decreased again after an initial increase, while the percentage of elective CS remained stable and high ever since.

first, more (relatively healthy) mothers had an elective CS. Second, physicians lowered their threshold with respect to performing an acute CS. These mechanisms lead to changes for both elected but also emergency CS probability after the TBT—more relatively healthy women were selected into both types of CS. As there are no differences in requirements regarding medical staff present for the two types of CS, we assume that the quality of the two types of CS is comparable. Admittedly, acute CS are most likely more stressful for the mother and they are performed if labor is complicated. At the same time, mothers who have an acute CS are the ones that have been selected for vaginal birth and are thus more likely to be comparable to the mothers who complete a natural delivery. The differences in the two types of CS and the mothers selected for them are discarded in our TOT analysis, which focuses on actual mode of delivery and does not distinguish the two types of CS.

In our analysis of the effect of CS on mother and child health, we consider the following outcomes: First, the APGAR score is a global measure of child health at birth. The APGAR score has been shown to predict infant mortality and future adverse outcomes, such as future health, cognitive ability and behavioral problems (Almond et al., 2005; Odd et al., 2008). The APGAR score is a summary score based on the evaluation of five criteria (Appearance, Pulse, Grimace, Activity, Respiration). We examine the probability of having a low score (<=7) at one and five minutes after birth. We additionally construct a measure of serious perinatal morbidity inspired by the measure used in the TBT. Our measure includes a set of perinatal infant health problems. We cannot perfectly match the TBT measure with our administrative data; nevertheless, we have a good proxy for whether the infant faces a serious perinatal problem related to mode of delivery.[28]

Furthermore, we consider different measures for child hospitalization after birth and in the first three and five years of life, one for more than three days and one for more than six days respectively. Similarly, we examine effects on the number of child outpatient visits at hospitals in the first three and five years of life. Due to data constraints, we analyze effects on the number of GP visits and visits at any specialized physician (e.g. ear or eye doctor) only for the first two years of the child's life.[29]

For maternal outcomes, we consider an indicator taking the value one for a set of maternal

[28]We use the ICD 10 diagnoses groups DP10-15, DP20-29, DP50-61, DP90-96, DO69, and APGAR<4 at five minutes (DVA00-DVA03).

[29]While not available to us at this stage, an analysis of children's prescription drug use could be used to examine conditions that do not require hospitalizations but treatment with drugs such as asthma.

complications after birth.[30] We do not examine maternal mortality at this stage, but we will consider it for future work.[31] While an analysis of future fertility and the probability of a consecutive CS for breech mothers would be very informative, our design does not allow this analysis.[32]

Table 1 shows summary statistics for all mothers with singleton children of parity higher than one in the period 1997 though 2006 and the subset of breech mothers. The table shows that mothers with breech pregnancies are not selected on observable characteristics such as maternal age at birth and maternal education. Also the percentage of mothers experiencing pregnancy complications that are unrelated to breech is similar in the two groups: relatively frequent conditions such as pre-eclampsia and diabetes have similar prevalence in both groups. Thus with respect to the characteristics considered here, the summary statistics confirm that breech pregnancies are as good as randomly distributed among mothers. While we do not exploit variation between the groups of pregnancies, the comparison in Table 1 illustrates differences in average outcomes for the two groups that are important to have in mind when interpreting our results. Table 1 also shows that the percentage of mothers having a CS is much higher for breech pregnancies. 72 percent of all breech mothers have a CS, while around 14 of all mothers have a CS. Breech mothers have a higher mean for both elective and acute CS in the period considered here.

Table 2 compares mothers and their breech babies born before and after the cut-off. CS rates increased considerably for the breech babies of interest from a mean of 62 percent to a mean of 79 percent. For a couple of background characteristics the table illustrates no changes, e.g. mothers age and education. We will examine potential changes in observables in greater deatail in subsection 4.2 on graphical analysis. The table indicates small decreases in the percentage of breech babies with a low APGAR score after the TBT. The changes in the outcome variables in the pre- and post-TBT period will also be subject to a more detailed analysis in subsection 4.2. For our hospitalization and outpatient contact outcomes we observe

[30] The ICD 10 codes included are DO85, DO860, DO861C, DO862A, DK556H, DO871, DO882D, DO702, DF53, DO990A, and operation codes KMWA, KMWB, KMWC, KKCH00, KJFA70, KJFA80, KLCD00, KMBA, KMBB, KMBC00, KTAB30. This coding follows the coding of maternal complications in The Danish National Board of Health (2005a).

[31] We will construct a variable for serious maternal complications including death as we have too few maternal deaths to separately examine this outcome.

[32] As we focus on mothers with higher parity breech babies, we run into very small samples with mothers having more than two children.

that while the post-TBT hospitalization means are in general lower than the pre-TBT means, outpatient contact means in the post-TBT period are higher. This finding suggest that there are trends in these outcomes that are important to control for in the following analysis.

4.2 Graphical Analysis

Figure 1 gives a first glance at our data. The figure shows the percentage of CS for all singleton pregnancies and for breech pregnancies over time. The graph is centred around the second half of the year 2000 and the means in the figure are for half-year intervals. The figure shows that while the CS rate is smooth for all singleton pregnancies, the probability of having a CS increases sharply from around 75 to around 83 percent for breech pregnancies after the second half of 2000.

This first look at aggregated data indicates a discontinuity in CS probability for breech babies around our cut-off, i.e., the DSOG meeting on December 4, 2000. To examine the importance of choice of bin width for our graphs of the first stage and the outcomes considered here, Appendix Table A.1 presents two tests—the "bin test" and "regression test" suggested in Lee and Lemieux (2010)—to inform bin width choice. The table shows the p-values for an F-test in two sets of regressions of outcomes on bin indicators: in the "bin test" we compare this model to an alternative one that includes a set of indicators for half the chosen bin width. In the "regression test" we compare the model to an alternative one that includes a set of interactions of the forcing variable with the bin dummies. Appendix Table A.1 shows the two tests for a bin width of 60 and 30 days respectively. The data window is then 20 and 40 bins on each side of the cut-off. In both tests we should not reject the null hypothesis for our bins to be acceptable. The table suggests that a bin width of around one month passes the tests for the outcomes that we consider, but also wider bins seem to be acceptable. In the following we generally focus on graphs with a bin width of 30 days, a bin size which represents the data well and secures enough observations per bin to avoid too much noisiness in the graphs.

All the following graphs plot means for non-overlapping bins to the left and right of the cut-off and impose a quadratic fitted line. If not noted differently, we use a data window of 40 bins on each side of the cut-off. This data window could appear to be large. However, considering the technologies of birth, we argue that the period of around three years on both sides of the cut-off is comparable with respect to the procedures used during pregnancy, such as attempted external cephalic version of breech babies, and during labor, such as CS.

Figures 2 to 4 show the percentage of CS for breech babies who are not first-born around the cut-off. Figure 2 uses our preferred bin widths (30 days) and the data window of 40 bins. Alternatively, Figure 3 shows a graph with narrower bins (15 days) and Figure 4 presents a graph for a smaller window of data (20 bins with 30 days each) for illustrative purposes.

All three figures for higher parity breech babies confirm a clear jump in the CS percentage irrespective of bin width and for different windows of data. In contrast, breech babies who are first-born do not experience the same discontinuous change (see Appendix Figure A.1). This difference is in accordance with the strict selection criteria already in place before the TBT. We do not argue that the TBT "invented" the idea of CS for breech babies. Rather than that it increased the strictness of evaluation and resulted in an even more careful selection of breech babies for vaginal delivery. Thus, in accordance with anecdotal evidence we find in our data that pre-existing and strong trends to select first-time mothers for CS use if their baby was in breech position were present. For first-time mothers, the TBT did not change physicians' practice. On the contrary, for higher parity mothers pre-TBT vaginal birth rates were high by international comparison and the change in physician practice was important. We therefore concentrate our analysis on this group.

Earlier studies on medical procedure use have focused on patients' characteristics and found that patients react on newly available information. Our application concentrates—as argued above—on changes on the part of physicians. Suggestive evidence for the validity of our argument is provided by Figure 5 that illustrates the jump in CS probability at the cut-off for mothers with different levels of education. The level of education could be an important determinant of changes in CS probability if—as suggested in models of health production—well-educated individuals are better to process information and act on it. This suggestion is supported by the evidence in Anderberg et al. (2011) and Price and Simon (2009). However, Figure 5 offers no support for the suggestion that mothers of high educational background drive the increase in the CS probability.[33] This finding is crucial because it gives us some sense about the marginal mother selected for CS after the cut-off. Rather than mothers' educational level, other factors such as mothers' overall health status and parity of the child

[33] We have also performed more formal falsification test for mother observables as outcomes: We have performed regressions for maternal education on the indicator for post-TBT birth, polynomials in the forcing variable, and other mother observables and find no indication for a discontinuous change in maternal education at the cut-off. Results are available on request.

most likely have determined this selection.

We continue with a graphical analysis of our outcome variables. To give suggestive evidence for the effect of the treatment, we plot our outcome variables in the same way as we have done for our first stage. Figures 6 and 7 show graphs for the outcomes APGAR of seven or below at one and five minutes, Figure 8 shows the percentage of children with serious perinatal morbidity. Figures 9 to 11 plot our hospitalization short-run child outcomes, namely the probability of hospitalization after birth and in the first three and five years of life (for more than six days respectively). Appendix Figures A.6 to A.8 show graphs for alternative hospitalization measures (more than 3 days of hospitalization) that lead to similar conclusions.

The figures indicate that the percentage of children with an APGAR score of seven or below decreases at the cut-off. For the five minutes APGAR score the figures suggest a significant decrease in the percentage of children with a low score after the cut-off. Also Figure 8 suggests a small decrease in the percentage of children with serious perinatal morbidity. Our figures for child hospitalizations do not indicate improvements or worsening of outcomes after the TBT.[34]

While all hospitalization figures show a downward trend over the period considered here, Figures 12 and 13 on children's outpatient visits show an increase in the percentage of children who have outpatient contacts to hospitals over the entire period. This trend seems to correspond to the concurrent decrease in the percentage of children experiencing a hospitalization in the period. We examine different measures for children's outpatient visits to hospitals in the first three and five years of their lives (see also Appendix Figures A.9 to A.11). Similarly to the graphs for hospitalizations we cannot detect any discontinuous changes around the cut-off.

While hospital stays and outpatient visits account for rather serious conditions and emergency room visits, it is possible that the health effects are more visible in measures that refer to less dramatic contacts with the medical system. Thus, Figures 14 and 15 show measures for the child's visits to GPs in the first two years of life. Figures 16 and 17 present measures of visits at specialized doctors outside hospitals (e.g., ear and eye specialists). For the percentage of children with above mean number of GP visits, the figure shows a jump at the

[34] A slight increase of birth hospitalizations shown in Appendix Figure A.6 could be driven by longer hospitalization of mothers after CS.

cut-off. After the cut-off the percentage of children with more than 17 GP visits in the first two years of life is lower. There is no indication of an increase in specialist visits. The GPs are gatekeepers in the Danish medical system and most children see a GP before they get referred to other physicians or hospitals.

In sum, our graphs indicate a strong and clear jump in the probability of having a CS around the cut-off. Furthermore, we see graphical evidence for positive effects of CS use for the probability of having a low APGAR score at five minutes and the probability of having above average number of GP visits. For other outcomes we do not detect clear discontinuities in our graphs.

Having established graphically that there is a jump in treatment probability around our cut-off, we now present further graphical evidence for the validity of our RD design. To make sure that manipulation of position at the cut-off is not an issue, Figure 18 plots the density of our forcing variable over a number of bins around the cut-off date. The figure shows that the percentage of breech babies per bin remains stable throughout the cut-off. Thus there is no indication that mothers or physicians manipulate the date of birth, an event, which we find highly unlikely to begin with.

If the TBT dissemination is the driving force behind the discontinuity in CS percentage for breech mothers, we should not observe increases in CS percentage for other groups of mothers. A simple falsification test is to look at the CS probability for another groups of mothers with high CS risk, such as mothers with pre-eclampsia. Their CS probability should not change discontinuously at the cut-off. Furthermore, at other dates—e.g. in the year before the dissemination of the TBT—we should not see similar discontinuities in CS probability for breech babies. Our graphical analysis shows that the percentage of mothers with pre-eclampsia having a CS does not change discontinuously at the cut-off. Furthermore, a placebo reform in 1999 does not result in a discontinuity in CS probability for breech mothers (see Appendix Figures A.12 and A.13).

Another potential concern in our analysis is that the TBT changed physicians' use of other technologies. Prime candidates for such parallel changes in technology are new diagnostic efforts to discover breech babies before labor and a higher success rate for attempted external cephalic versions. Midwives and physicians could pay more attention to identifying breech babies before labor and be more successful with their attempts to turn the baby. Or, as an alternative, physicians could "bother less" and schedule more CS without any preceding attempt to perform an external cephalic version as a consequence of the TBT.

Figure 19 shows the percentage of breech babies that are diagnosed before labor. The figure shows a small increase in the percentage of pre-labor diagnosed breech babies after the cut-off but the variation in the percentage of discovered breech babies is big across bins. Having discovered the breech position allows the physician to do two things: first, try an external cephalic version of the baby and thereby try to prevent a breech birth, and second, plan for the delivery of the baby, i.e., schedule an elective CS or attempt a vaginal birth.

Figure 20 shows the percentage of all attempted—successful and unsuccessful—external cephalic versions among all breech babies.[35] As the figure shows, the percentage of breech babies with an attempted external cephalic version does not change discontinuously at the cut-off. Furthermore, as the fitted line illustrates, the percentage of breech babies with attempts remains rather stable over time. These two observations do not suggest that physicians tried more or less hard to prevent breech births. Additionally, Figures 21 and 22 show the percentage of successful and unsuccessful external cephalic versions. These figures do not suggest a higher rate of success after the cut-off.[36] Thus the changes in pre-labor diagnoses and successful external cephalic versions are minimal when compared to the size and significance of the jump in CS probability. This finding indicates that no important parallel changes in this technology coincide with the increase of CS use for breech babies.[37]

Another concern with respect to the potential manipulation of treatment status results from the mothers' option to change hospitals during their pregnancy. Mothers could change hospitals during pregnancy and could pick hospitals according to their CS policy. Figure 23 shows the percentage of breech mothers who change their hospital between first trimester prenatal care visit and birth.[38] As the figure shows, the percentage of mothers who change their hospital is very small, below ten percent of breech mothers. Moreover and importantly, the percentage of changers remains unchanged after the cut-off, i.e. there is no indication that mothers change hospitals more often than before the TBT.

[35] ICD 10 codes KMAB10 and KMAB20.

[36] The percentages shown here are relative to all breech babies. Of all attempted external cephalic versions, around 50 percent are successful.

[37] As mentioned in section 2.2, changes in the standard use of ultrasound do not occur either.

[38] We assign the fist hospital to mothers by using the earliest ICD 10 codes DZ34 and DZ35 during pregnancy (earliest prenatal care visit). We use the mother's birth diagnosis to assign the birth hospital. We find identical results when using the hospital with the child's first hospitalization as birth hospital.

Finally, the continuity assumption requires other maternal characteristics to be smooth at the cut-off. Figure 24 shows that the jump in CS probability after the cut-off is not driven by mothers who have had a previous CS. Around one third of all breech mothers have had a previous CS and this percentage remains close to unchanged throughout the period. Appendix Figure A.14 shows that there is no discontinuity for a set of other observable background characteristics—mothers' age at birth, mother's educational attainment and immigrant status, and child's sex.[39] Taken together, we believe that our assumption of an exogenous change in treatment probability at the cut-off date is justified.

5 Results

This section presents our results for the effect of CS on child and mother health outcomes. We present benchmark estimates that can be compared to Tharin et al. (2011), furthermore we present reduced form estimates for the effect of the TBT-dissemination on outcomes, and estimates for the marginal breech birth (RD estimates instrumenting for CS). We present results for our parametric specifications and for local linear specifications.

The study of Tharin et al. (2011)—that uses Danish data and examines the effects of the TBT dissemination on infant outcomes—is a natural starting point to put our analysis for infant short-run health outcomes into perspective. While they focus on the planned mode of delivery and on ITT effects—as discussed earlier—their risk ratios can serve as a point of reference for our results. As they use a large data window around the TBT, a comparison of their results and our RD results can inform about the impact of data range used for the results obtained. Furthermore, we argue that our fuzzy RD estimates are relevant for the marginal mother and child who change mode of delivery due to the TBT dissemination. This effect is of obvious policy importance.

We look at the one similar outcome that both this study and Tharin et al. use, namely APGAR score at five minutes. As opposed to their study, we use a shorter data period (they use 2 additional post-TBT years which we do not have in our data at this point), we exclude first parity children from our analysis, and we examine the probability of having an APGAR score below seven (instead of six).[40] The risk of having a low APGAR score decreases from

[39]Finally, we also conducted a RD analysis using our controls as outcome variables and found no indication of a discontinuous change.

[40]Tharin et al. (2011) use the measure APGAR at five minutes smaller or equal six. In our future work will

the pre- to the post-TBT period. Tharin et al. (2011) report a significant reduction of the risk of a low APGAR score at five minutes (APGAR<=6) in their data (RR 0.83, CI 0.73-0.95). Calculating a risk ratio for having an APGAR score below seven at five minutes and doing so by using all our available data, we find a risk ratio of 0.74 (CI 0.5-1.1, p=0.14). However, reducing the data window to around one year on both sides of the TBT dissemination results in a significant and very large reduction in the incidents of APGAR<=7 in our data (Risk Ratio 0.26, CI 0.09-0.79, p=0.01). While we have to keep in mind that the low APGAR group is a very small group of children, this risk ratio can be interpreted as a 74 percent decrease in the risk of a low APGAR score for the exposed group (after TBT). These ratios do not take the changing composition of CS breech babies into account and the range of data used appears to be of great importance. In the following analyses we focus on our RD estimates that—as argued above—emphasize the importance of choice of data window for the conclusions made and help to identify the effect of CS on the marginal child, whose mode of delivery was affected by the dissemination of the TBT results.

As a first step, Table 3 presents estimates for the change in outcomes at the cut-off (the reduced form effect). Both columns include quadratic polynomials in the forcing variable and an interaction of the pos-TBT indicator with the forcing variable. Furthermore, column two includes controls for mother observables (consult table notes). The estimates are not sensitive to the inclusion of these controls. We find that the effect on the probability of having a low APGAR score at five minutes decreases by about 2.1 percentage points. This is a very large effect and in line with the above calculated risk ratio in a narrow window around TBT dissemination. For all other outcomes—except out patient visits in the first 5 years of life—we find no significant effects.

The presented graphical evidence clearly indicates a jump in treatment probability (CS probability) around the TBT cut-off that can be used to identify the effect of CS on the marginal breech baby. In addition, Table 4 shows estimates of the jump from local linear regressions. Table 4 indicates that from a bandwidth of around 180 days on each side of the cut-off the estimate for the jump in treatment probability stabilizes between 0.14 and 0.15 and is significant in our local linear regression (bootstrapped standard errors, number of replications 200). This finding is robust to kernel choice. Appendix Figures A.2 to A.5

harmonize our analysis, sample and outcome measures as much as possible with their calculations and provide a better and more detailed comparison of findings and their implications.

illustrate the estimates of our local linear regressions for different bandwidths. For varying bandwidths the graphs suggest that CS probability can be described by a linear function on both sides of the cut-off and that the jump at the cut-off is not driven by functional form assumptions. For our main analysis this finding suggests that second order polynomials should be more than flexible enough to model the first stage.

Table 5 presents estimates for our first stage regression for the CS indicator including different order polynomials and controls. The table shows that across specifications being born after the cut-off increases CS probability. In our preferred specification (quadratic) this increase is about 15 percentage points and very similar to the estimate of the jump from the local linear regression. First stage F-values for our treatment indicator are also reported in Table 5 and confirm the relevance of our instrument. However, in the specifications based on very tight windows of data—and rather few observations—F-values are only around 7, indicating a potential weak instrument problem. We thus concentrate on the models including more data around the cut-off.

Turning to our instrumental variable estimation, Table 6 presents the results for the probability of having a low APGAR score at one and five minutes and our measure of child morbidity. Table 7 presents results for the probability of being hospitalized and Table 8 presents results for the probability of having outpatient visits. Table 9 presents results for GP and specialist visits. For mother outcomes, Table 10 shows the results for the probability of experiencing infections and a number of post-birth complications.

In all tables, each cell presents the results of a separate regression. Each cell presents the respective estimate for the treatment indicator and robust standard errors in parentheses. Columns one and two show our IV estimates for models including quadratic polynomials, quadratic splines (an interaction of the treatment indicator with the forcing variable to allow for different slopes on both sides of the cut-off), and with and without additional control variables for the entire range of data from 1997 through 2006. Columns three and four estimate the same models including cubic polynomials. While columns one to four use all available data, columns five and six estimate the quadratic model with a data window of around 40 months on both sides of the cut-off. The final two columns estimate local linear regressions with a rectangular kernel and the rule of thumb (RoT) bandwidth and double this bandwidth.

Overall, we find no robust indication for persistent health benefits or costs of CS for the marginal CS child and her mother. The one exception is the probability of having a low APGAR score at five minutes. Panel one and two of Table 6 show our results for the probability of a low APGAR score at one and five minutes, respectively. While the estimates for the APGAR score at one minute are negative and thus indicate a decrease of the probability of having a low score at one minute, they are not significantly estimated. However, our estimates for the five minute APGAR score are similar in size and show across specifications that children born by CS have a significantly lower probability of having a low score at this point. The lack of significance at one minute could be due to greater measurement error in this very early measure.

At five minutes the APGAR score is more reliable and most likely a better indicator of child health at birth. We find very large effects on this outcome measure. The marginal breech babies born by CS are 14 to 15 percentage points less likely to have low APGAR score at five minutes. This estimate indicates a large effect of CS on the marginal child's probability of having a low APGAR score and this effect is robust across specification. We also find similar results in the local linear regressions. We find no significant results for our measure of serious child morbidity.

In line with the reduced form estimates, Table 7 and Table 8 show that there are no significantly estimated effects of CS for the marginal breech baby with respect to different hospitalization measures and outpatient visits. Additionally, the measures that include a longer period and measure longer hospitalizations/more outpatient visits tend to be smaller in absolute size. This finding indicates that for the measures that capture more severe health issues differences between the marginal CS children and children born by vaginal delivery are—if at all present—very small. The same conclusion holds for the results for our outpatient contact measures presented in Table 8.

While Table 9 shows large point estimates for the effect of CS on the probability of having more than the average number of GP visits in the first two years of life for the marginal breech baby, they are only significantly estimated in the cubic model. This finding could suggest some moderate health improvements for CS children over and above birth outcomes. The GP is the most important contact of children to the health care system and fewer GP contacts could indicate a decrease in relatively mild health incidents, as these are the ones most likely treated by GPs. Finally, results for specialist visits indicate no health effects of CS for the marginal breech baby.

For maternal short-run outcomes, Figure 10 shows very small and insignificantly estimated coefficients. This finding is not surprising and mirrors the graphical evidence which does not indicate short-run health benefits or risks for post-birth complications for the marginal CS mother.

The final two columns of Tables 6 to 10 present estimates that are based on local linear regressions and a comparison to our preferred 2SLS estimates (quadratic polynomials) shows that for the outcome for which we find significant results—APGAR score at five minutes—the estimates compare well. Figures 25 to 28 plot the result from our IV regressions with a quadratic polynomial against different local linear regression estimates and their confidence bands. For our APGAR results, child morbidity and above average GP visits, the local linear estimates are very similar to our 2SLS estimate that is based on all available data. Thus this comparison confirms that our results are not driven by functional form assumptions.

The lack of precision in our estimates could be a problem that relates to the sample size required to detect the effect of CS in a RD design. Appendix Figures A.15 to A.20 show the coefficients and confidence bands from our 2SLS estimations for different bandwidths of data for the APGAR, morbidity, GP visits, and the hospitalization outcomes. The size of the estimated effects is rather stable and not particularly sensitive to the data window chosen. However, most of the estimated effects are not significantly different from zero although confidence bands generally tighten when using more data (larger data windows).

6 Discussion and Conclusion

This paper exploited a unique exogenous variation in CS probability for a well-defined subgroup of all pregnancies—breech pregnancies at term—to examine the effect of CS on mother and child health for the marginal mother, i.e., the mother who would have had a vaginal birth before the TBT but has a CS after the TBT. This complying mother—and her baby—are most likely relatively healthy. In accordance with this suggestion, we find no across-the-board health benefits for the baby and no short-run benefits or risks for the mother. We find large estimates when examining one child health outcomes, namely the probability of having a low APGAR score at five minutes. Furthermore, there is weak evidence for the marginal CS baby having a lower probability to visit her GP very often in the first two years of life. We find no significant effects for a range of health measures including perinatal morbidity, child hospitalization and outpatient visits, as well as for mothers' birth-related health problems. These findings suggest that in a strict regime for breech births like in Denmark physicians have done a good job already before the TBT in selecting mothers for their mode of delivery. Thus health benefits from a stricter regime are modest and seem to be immediate, i.e., restricted to the time directly after birth. At the same time, we can conclude that for the health outcomes considered, we find no negative effects, i.e. CS for the marginal child does not seem to harm immediate health outcomes.

However, future research should also consider other potential health benefits measured, e.g., as specific diagnoses of children. Although being a strong evaluation method and offering "local randomization", RD designs require more data than randomized controlled trials to detect similar effects at a given confidence level. Thus part of our not finding significant results for rather rare conditions (e.g. for mothers) could be due to power issues.

As a consequence of the TBT, CS rates have increased significantly in many countries, among them Denmark, and this development has heated a debate on whether physicians today are geared to perform vaginal deliveries with breech babies (Turner, 2006). This knowledge will also in the future be important as some breech babies will be detected too late to perform a CS (among them, e.g., twins where the second baby faces a higher breech risk). Thus the unpreventable vaginal breech births might today encounter elevated health risks as a consequence of the TBT and its impact on physician practice. The increasing CS rates for relatively healthy breech mothers might thus have side effects for mothers with breech babies that have to be delivered naturally.

Our paper illustrates that physicians' practice can change rapidly due to the availability of new information published in highly ranked scientific journals such as the Lancet. This finding poses the question which mechanisms drive this change - physicians' preferences, liability concerns, and peer effects are some of the prime candidates to be studied in this context. Moreover, our paper shows that supply-side factors—such as physicians' behavior—are a non-trivial determinant of increased procedure use for childbirth. While recent public debates focus on the rise of CS on maternal request, our paper shows that increased procedure use for the marginal and comparably healthy woman is determined greatly by physicians' behavior.

Given that the focus of this paper is on a specific subgroup of babies—breech at term—the question remains as to whether our results are generalizable for other patient groups. Additionally, the external validity of the RD design—which focuses on data close to a clearly defined cut-off—and the fuzzy RD design—which only estimates the average treatment effect for compliers at the cut-off—remains restricted. The marginal mothers are clearly different from mothers who always have a CS. However, they are most likely a very interesting group from a policy perspective: The marginal mothers are relatively healthy and so are their children. Our analysis suggests that there are modest health benefits for the marginal CS baby. While we do not find elevated risks for the marginal mother, we have not examined other important aspects—longer-run health risks for mothers and potential problems in consecutive pregnancies. Further research is warranted to shed light on the costs and benefits of CS use.

References

Aalborg Hospital, "Guidelines for births with singleton breech at term, Aalborg Hospital," Guideline 2001.

Almond, D, KY Chay, and DS Lee, "The Costs of Low Birth Weight," *The Quarterly Journal of Economics*, 2005, *120* (3), 1031–1083.

Anderberg, D, A Chevalier, and J Wadsworth, "Anatomy of a health scare: Education, income and the MMR controversy in the UK," *Journal of Health Economics*, 2011, *30* (3), 515–530.

Angrist, J and JS Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2008.

Baicker, K, KS Buckles, and A Chandra, "Geographic Variation In The Appropriate Use Of Cesarean Delivery," *Health Affairs*, 2006, *25* (5), 355–367.

Bewley, S and A Shennan, "Peer Review and the Term Breech Trial," *The Lancet*, 2006, *369* (9565), 906.

Carayol, M, Bl Blonde, J Zeitlin, G Breart, and F Goffinet, "Changes in the rates of caesarean delivery before labour for breech presentation at term in France: 1972–2003," *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 2007, *132* (1), 20 – 26.

Clausen, JA, "Causes for changes in the method of delivery. Why was the TBT so influential? [Begrundelser for forandring af fødselspraksis. Hvorfra får TBT studiet sin autoritet?]," Masters Thesis, Masteruddannelse i Humanistisk Sundhedsvidenskab og Praksisudvikling, Åben Uddannelse, Århus Universitet 2003.

Currie, J and WB MacLeod, "First Do No Harm? Tort Reform and Birth Outcomes," *The Quarterly Journal of Economics*, 2008, *123* (2), 795–830.

Danish Society of Obstetrics and Gynaecology, "Guidelines for births with singleton breech at term," Guideline, Danish Society of Obstetrics and Gynaecology 1998.

_ , “[DSOG Newsletter 4/2001],” Newsletter, Danish Society of Obstetrics and Gynaecology 2001.

Declercq, E, F Menacker, and M MacDorman, “Maternal Risk Profiles and the Primary Cesarean Rate in the United States, 1991-2002,” *Am J Public Health*, 2006, *96* (5), 867–872.

Del Bono, E, M Francesconi, and N Best, “Health Information and Health Outcomes: An Application of the Regression Discontinuity Design to the 1995 UK Contraceptive Pill Scare Case,” Economics Discussion Papers 696, University of Essex, Department of Economics June 2011.

Epstein, A and S Nicholson, “The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections,” Working Paper 11549, National Bureau of Economic Research August 2005.

Fan, J and I Gijbels, *Local Polynomial Modelling and Its Application*, Chapman and Hall, London, 1996.

Glezerman, M, “Five years to the term breech trial: The rise and fall of a randomized controlled trial,” *American Journal of Obstetrics and Gynecology*, 2006, *194* (1), 20 – 25.

Goffinet, F, M Carayo, JM Foidart, S Alexander, S Uzan, D Subtil, and G Bréart, “Is planned vaginal delivery for breech presentation at term still an option? Results of an observational prospective survey in France and Belgium,” *American Journal of Obstetrics and Gynecology*, 2006, *194* (4), 1002 – 1011.

Grant, D, “Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project,” *Journal of Health Economics*, January 2009, *28* (1), 244–250.

Gregory, KD, SC Curtin, SM Taffel, and FC Notzon, “Changes in indications for cesarean delivery: United States, 1985 and 1994,” *Am J Public Health*, 1998, *88* (9), 1384–1387.

Gruber, J and M Owings, “Physician Financial Incentives and Cesarean Section Delivery,” *Rand Journal of Economics*, 1996, *27* (1), 99–123.

— , **J Kim, and D Mayzlin**, "Physician fees and procedure intensity: the case of cesarean delivery," *Journal of Health Economics*, 1999, *18* (4), 473 – 490.

Hannah, ME, WJ Hannah, SA Hewson, ED Hodnett, Sl Saiga, and AR Willan, "Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial," *The Lancet*, 2000, *356* (9239), 1375 – 1383.

Hofmeyr, GJ, ME Hannah, and TA Lawrie, "Planned caesarean section for term breech delivery," *Cochrane Database of Systematic Reviews 2003, Issue 2*, 2003.

Imbens, G and T Lemieux, "Regression Discontinuity Designs: A Guide to Practice," NBER Working Papers 13039, National Bureau of Economic Research, Inc 2007.

Kotaska, A, "Inappropriate use of randomised trials to evaluate complex phenomena: case study of vaginal breech delivery," *BMJ*, 10 2004, *329* (7473), 1039–1042.

Lee, DS and T Lemieux, "Regression Discontinuity Designs in Economics," Working Paper 14723, National Bureau of Economic Research February 2009.

— **and** — , "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

Ministry of the Interior, "Activity-based reimbursement in the medical sector, January 2003 [Takststyring på sygehusområdet, januar 2003]," Report 2003.

Odd, D E, F Rasmussen, D Gunnell, G Lewis, and A Whitelaw, "A cohort study of low Apgar scores and cognitive outcomes," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 2008, *93* (2), F115–F120.

Phipps, H, CL Roberts, N Nassar, CH Raynes-Greenow, B Peat, and EK Hutton, "The management of breech pregnancies in Australia and New Zealand," *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 2003, *43* (4), 294–297.

Price, J and K Simon, "Patient education and the impact of new medical research," *Journal of Health Economics*, 2009, *28* (6), 1166 – 1174.

Rietberg, CT, PM Elferink-Stinkens, and GHA Visser, “"The effect of the Term Breech Trial on medical intervention behaviour and neonatal outcome in The Netherlands:an analysis of 35,453 term breech infants",” *British Journal of Obstetrics and Gynaecology*, 2005, (112), 205–209.

Rochut, J, “Health Care Supply, Payment System and Medical Practice: Evidence from Obstetric Practice,” PhD Thesis, University of Lausanne 2010.

Shearer, EL, “Cesarean section: Medical benefits and costs,” *Social Science and Medicine*, 1993, *37* (10), 1223 – 1231.

Tharin, JEH, S Rasmussen, and L Krebs, “Consequences of the Term Breech Trial in Denmark,” *Acta Obstetricia et Gynecologica Scandinavica*, 2011, *90* (7), 767–771.

The Danish National Board of Health, “Ceasarean Section 1973-2005 [Kejsersnit 1973-2005],” Report, The Danish National Board of Health 2005.

_ , “Ceasarean Section on maternal request - a medical assessment [Kejsersnit på moders ønske. En medicinsk teknologivurdering],” Report, The Danish National Board of Health 2005a.

Triunfo, P and M Rossi, “The effect of physicians’ remuneration system on the Caesarean section rate: the Uruguayan case,” *International Journal of Health Care Finance and Economics*, 2009, *9*, 333–345. 10.1007/s10754-008-9054-y.

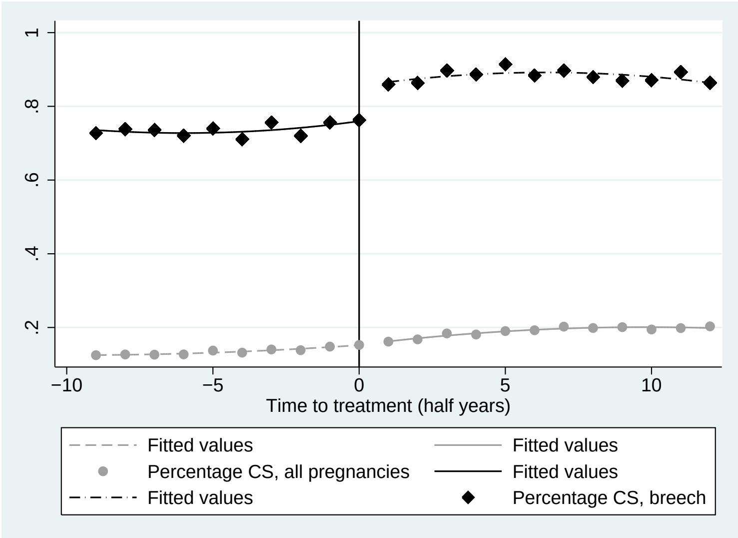
Turner, MJ, “The Term Breech Trial: Are the clinical guidelines justified by the evidence?,” *Journal of Obstetrics and Gynecology*, 2006, *26* (6), 491–494.

van der Klaauw, WH, “Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics,” *LABOUR*, 2008, *22* (2), 219–245.

Vartan, CK, “Cause of breech presentation,” *The Lancet*, 1940, *235* (6083), 595–596.

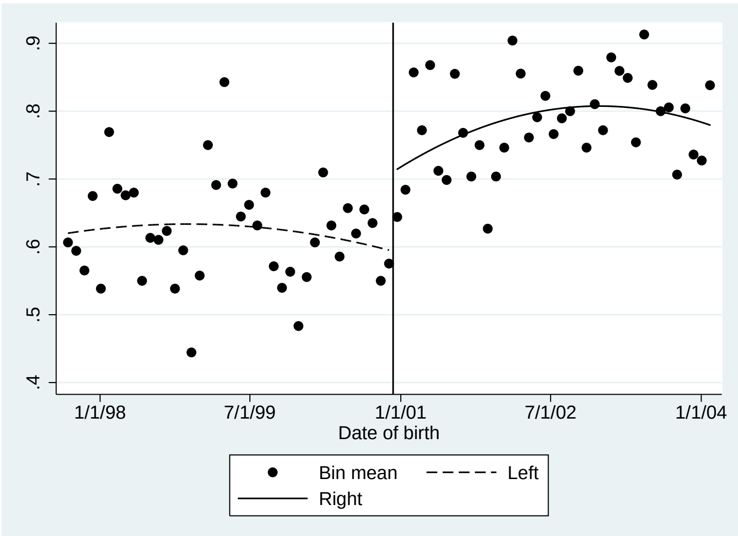
Zarko, A, D Declan, and Gyte GML, “Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour,” *Cochrane Database of Systematic Reviews 2006, Issue 3*, 2006.

Figure 1: C-section rate for all singletons and breech pregnancies 1996-2006



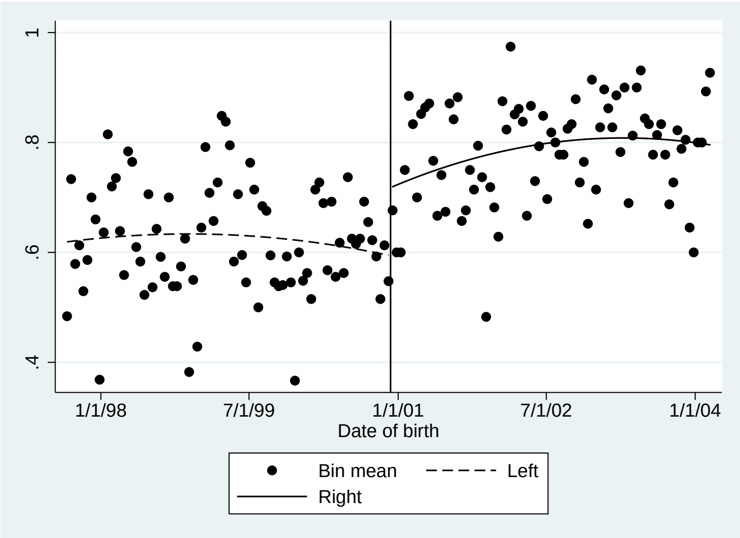
Notes: All singletons irrespective parity. Means per half calendar year.

Figure 2: C-section rate for breech pregnancies, parity>1



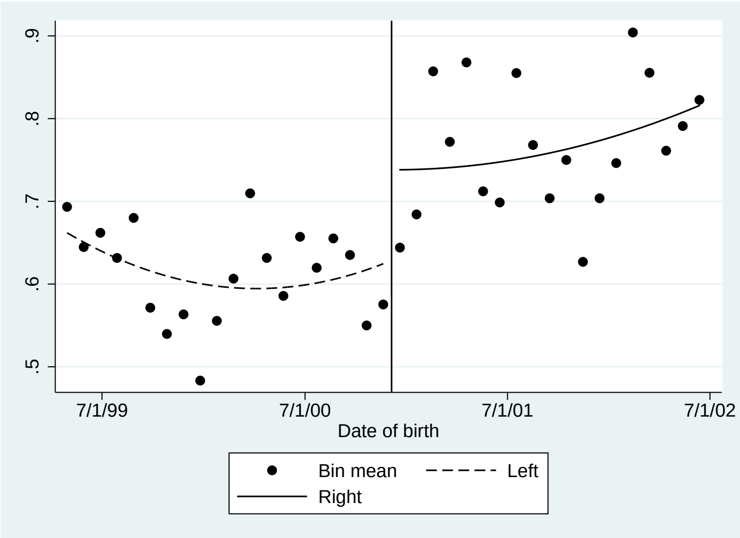
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 3: C-section rate for breech pregnancies, parity>1



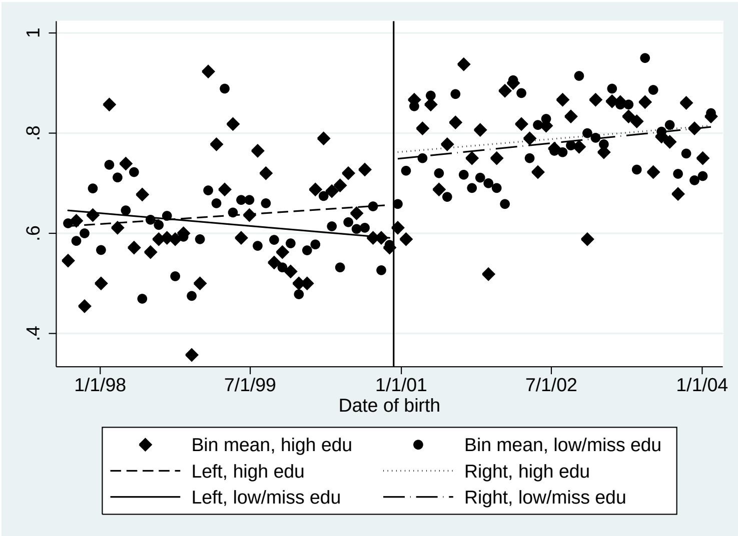
Notes: Width of bins: 15 days, 80 bins on each side of the cut-off.

Figure 4: C-section rate for breech pregnancies, parity>1



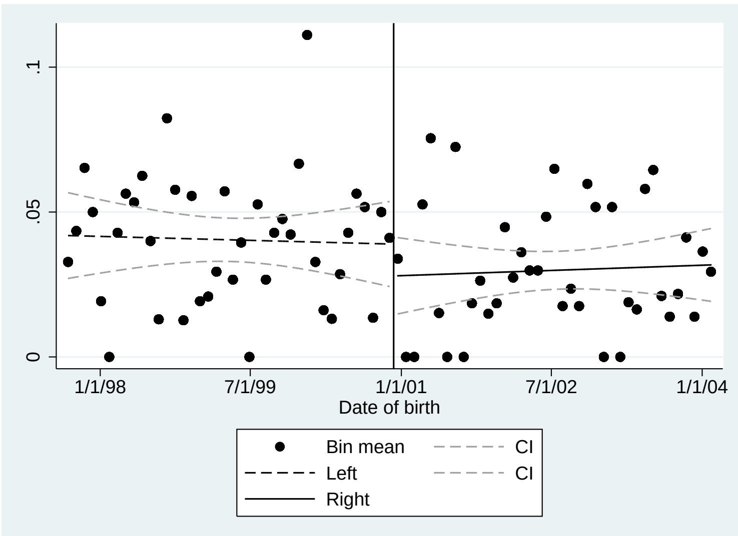
Notes: Width of bins: 30 days, 20 bins on each side of the cut-off.

Figure 5: C-section rate for breech pregnancies by mother's education (mothers with university degree vs all other mothers), parity>1



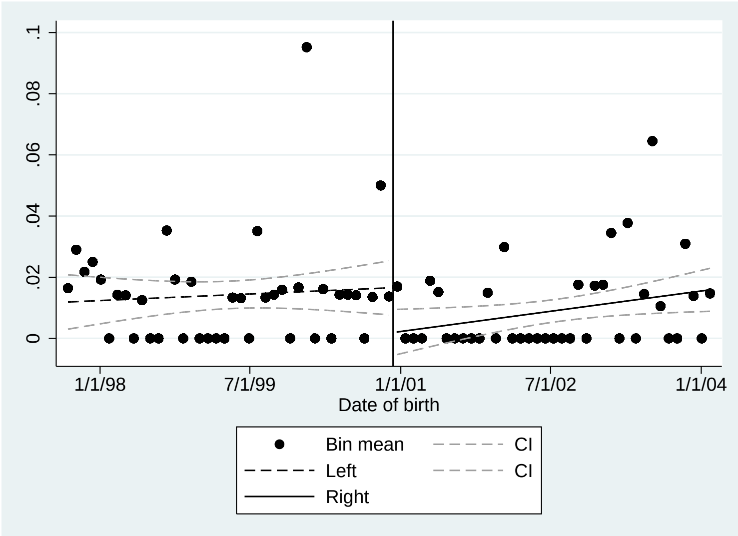
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 6: Probability of APGAR score <=7 at 1 min for breech babies with parity>1



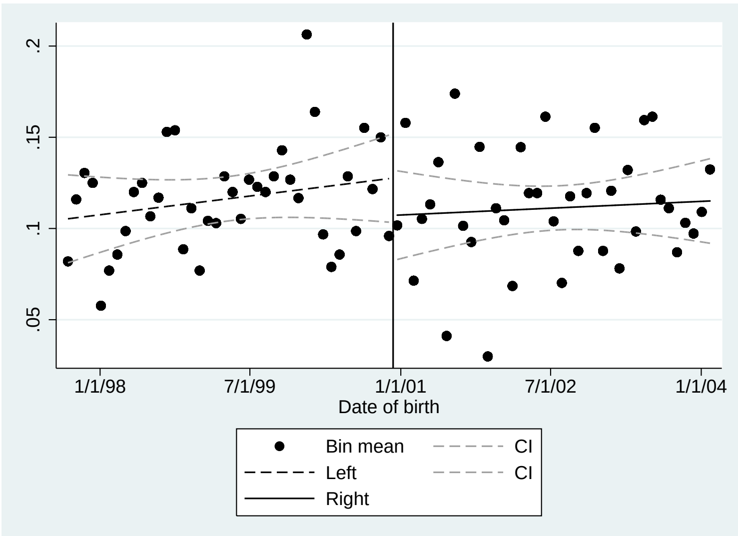
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 7: Probability of APGAR score <=7 at 5 min for breech babies with parity>1



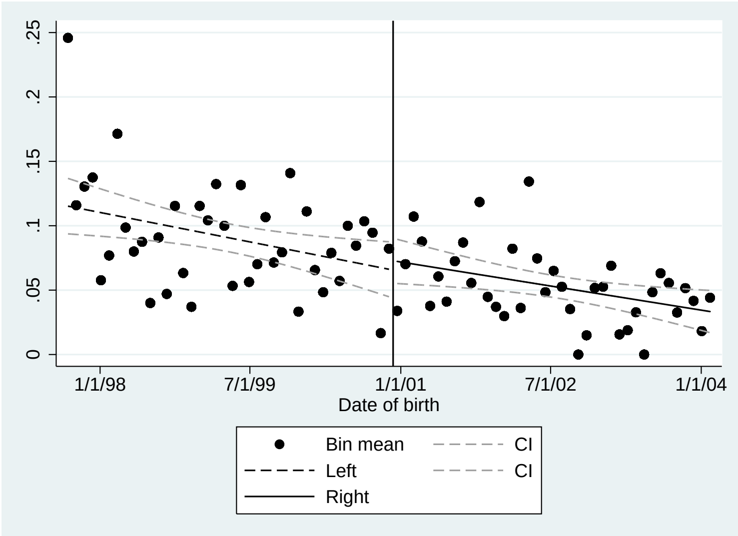
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 8: Serious perinatal morbidity for breech babies with parity>1



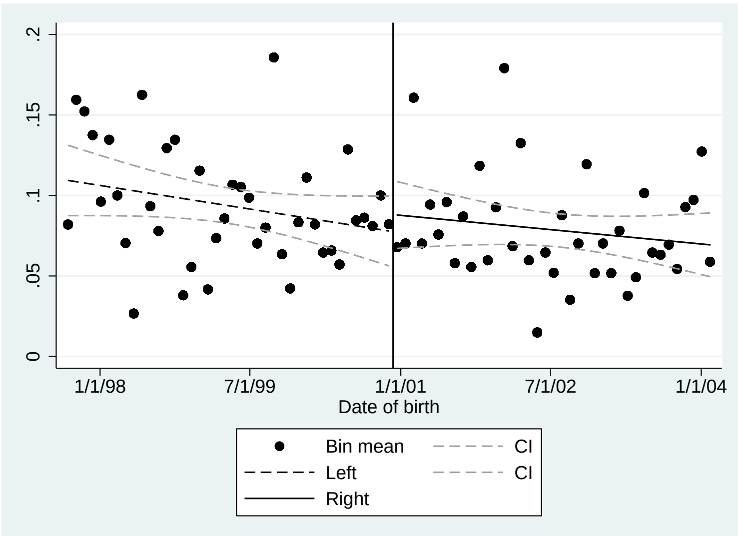
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 9: Percentage of children with more than 6 days of hospitalization at birth



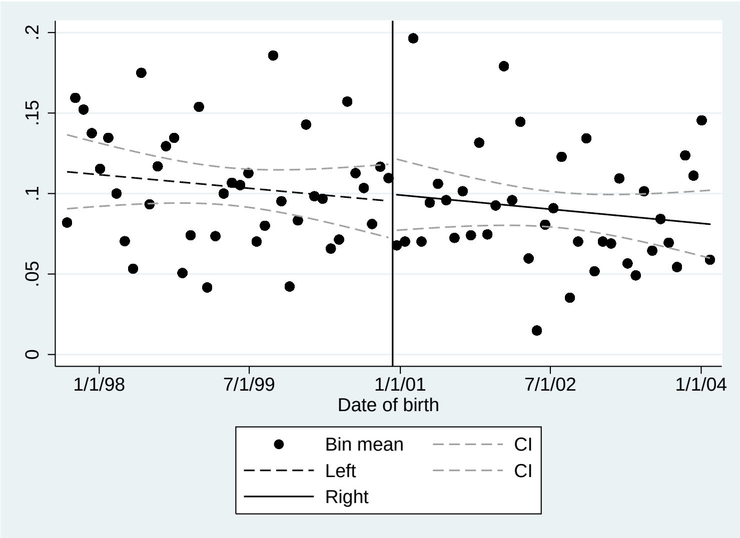
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 10: Percentage of children with more than 6 days of hospitalization in first 3 years



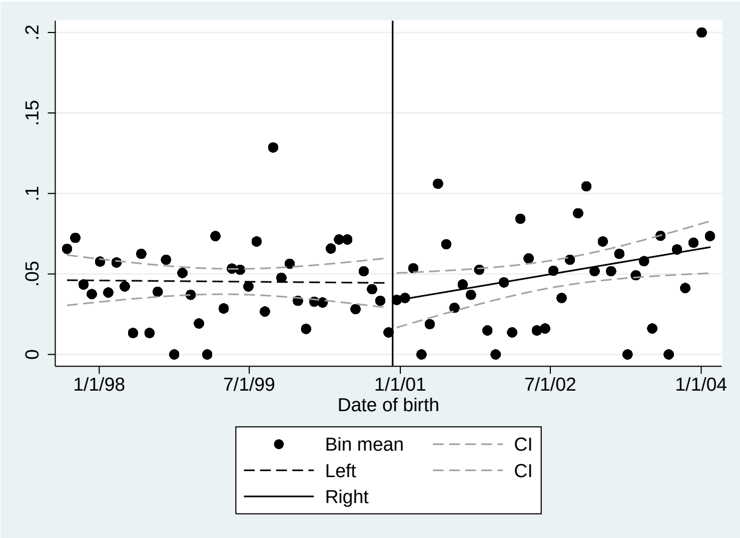
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 11: Percentage of children with more than 6 days of hospitalization in first 5 years



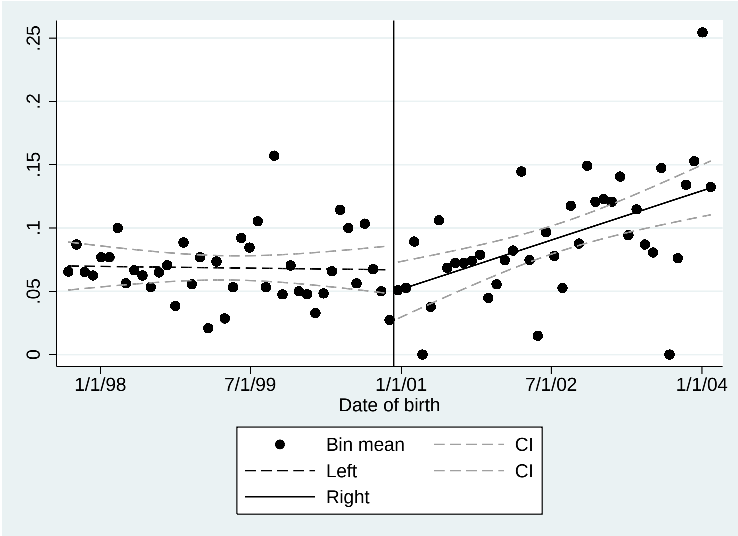
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 12: Percentage of children with more than 5 outpatient visits in first 3 years



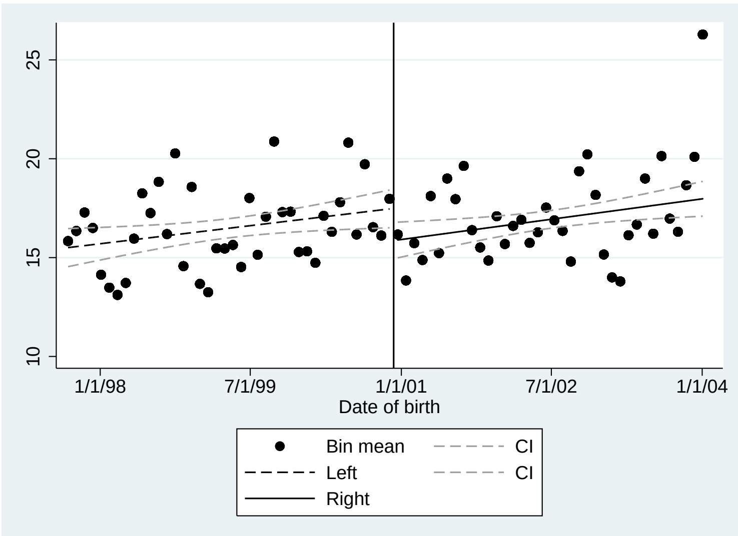
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 13: Percentage of children with more than 5 outpatient visits in first 5 years



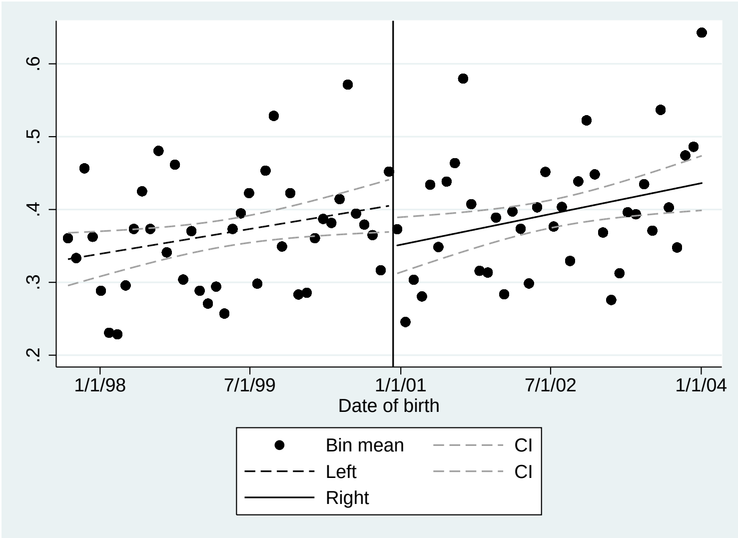
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 14: Mean nbr of GP visits in first 2 years



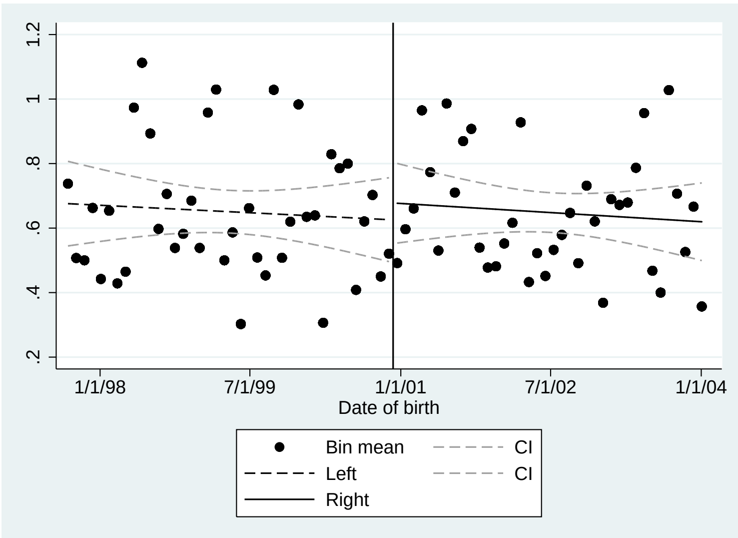
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 15: Percentage of children with more than 17 (mean) GP visits in first 2 years



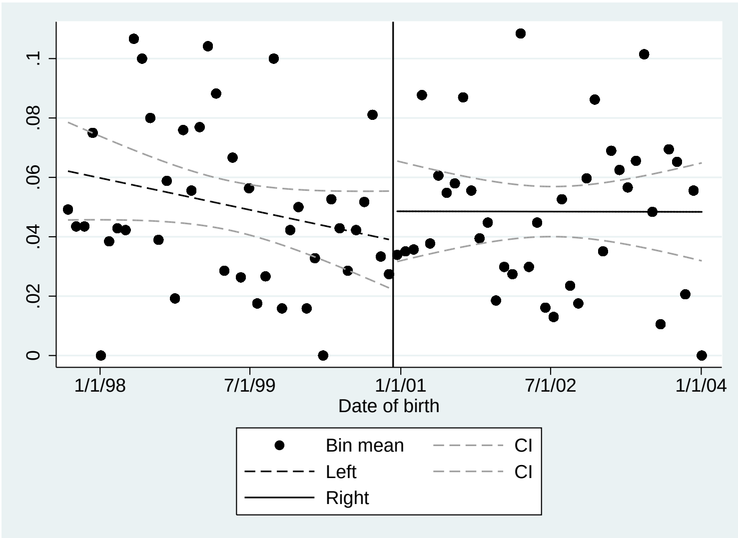
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 16: Mean nbr of specialist visits in first 2 years



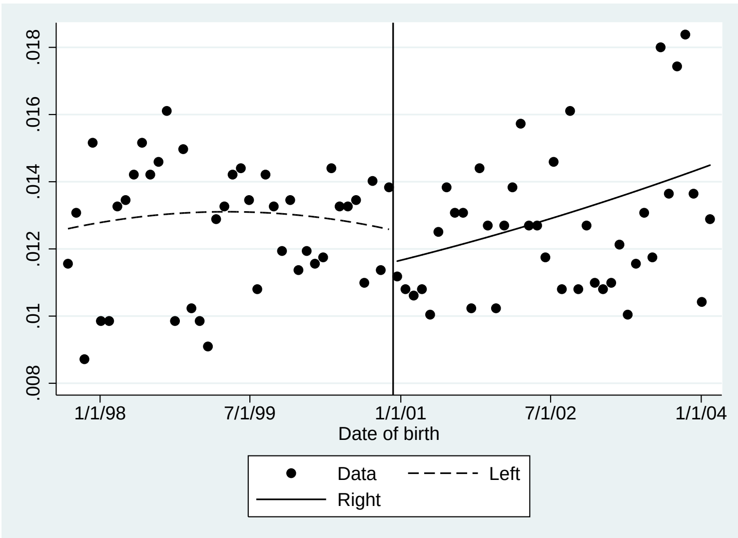
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 17: Percentage of children with more than 3 (mean:0.65) specialist visits in first 2 years



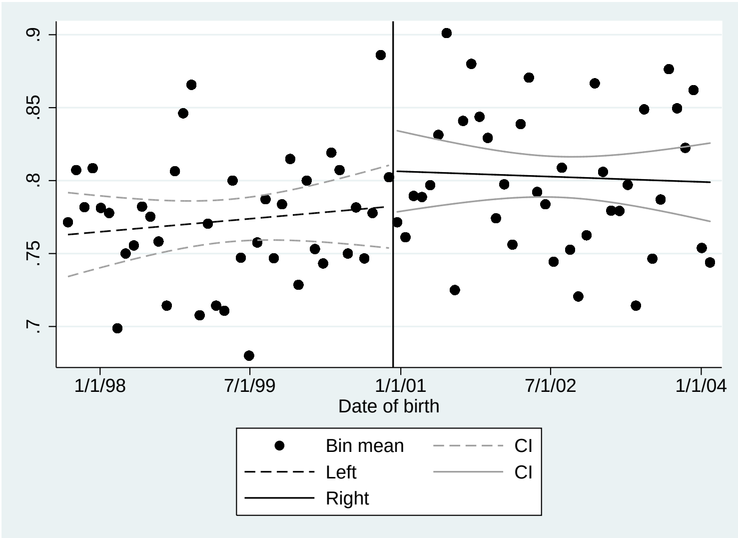
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 18: Density of the forcing variable



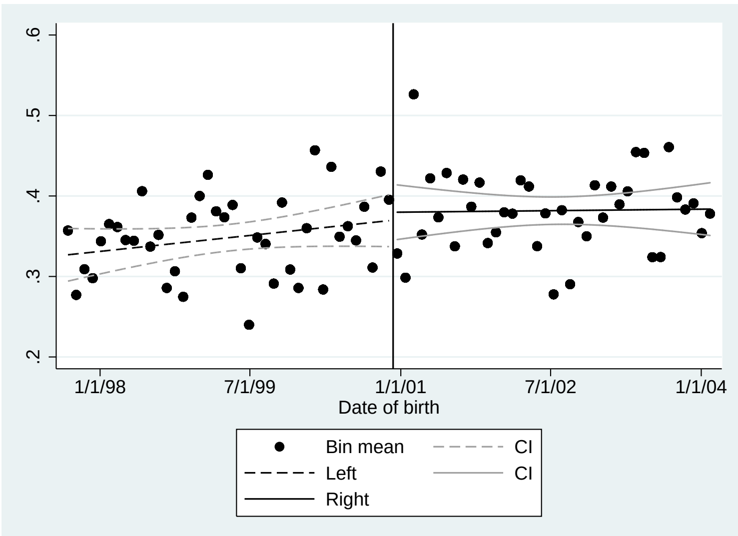
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 19: Test for changes in other technologies: Percentage of breech babies diagnosed before labor



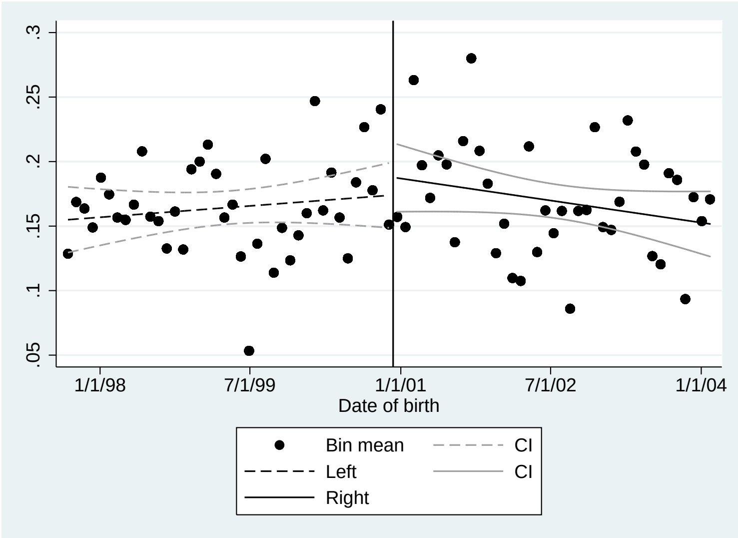
Notes: Diagnose for breech is given at least one day before birth date of the child. Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 20: Test for changes in other technologies: Percentage of all breech babies with attempted or successful external cephalic version



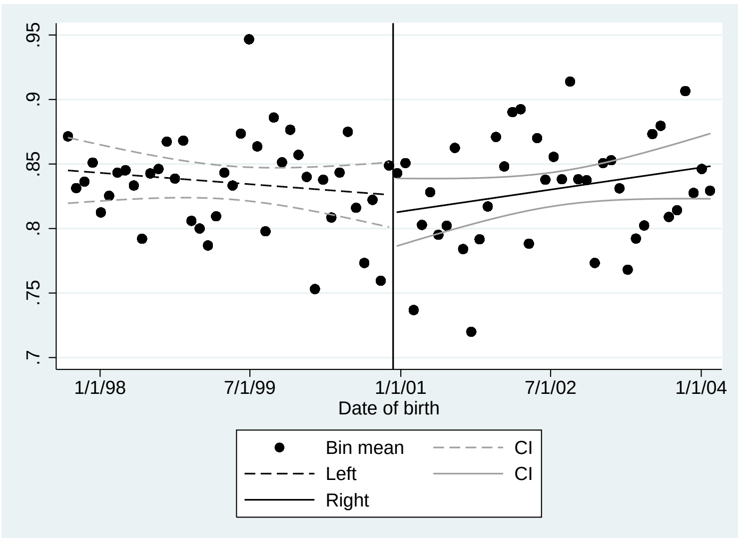
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 21: Test for changes in other technologies: Percentage of all breech babies with successful external cephalic version



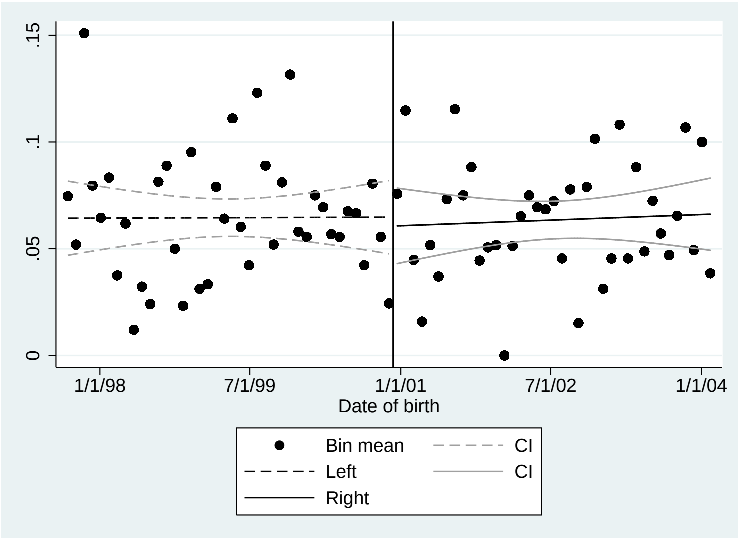
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 22: Test for changes in other technologies: Percentage of breech babies without successful external cephalic version



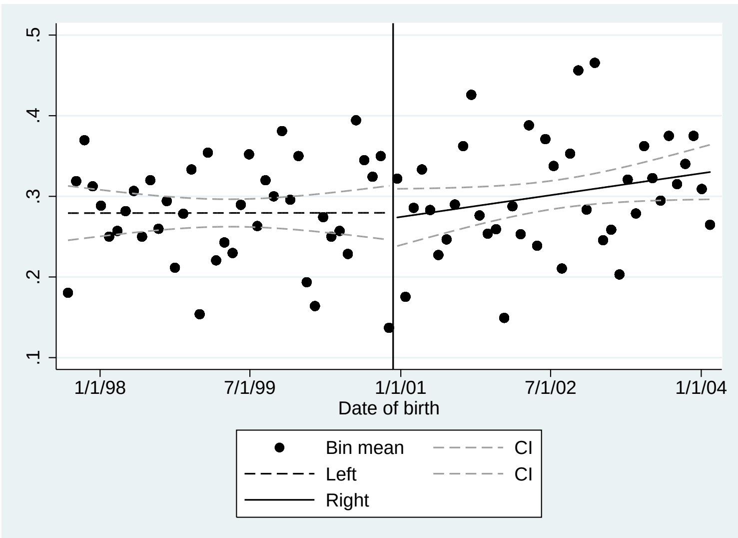
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 23: Test for other changes: Percentage of breech mothers who change their hospital during pregnancy



Notes: Width of bins: 30 days, 40 bins on each side of the cut-off. Percentage of mothers with different hospitals for first pregnancy visit and birth.

Figure 24: Test for jump in observable characteristics: Previous CS



Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure 25: Local linear regression estimates for varying bandwidth and estimate from quadratic model for APGAR<=7 at 1 min

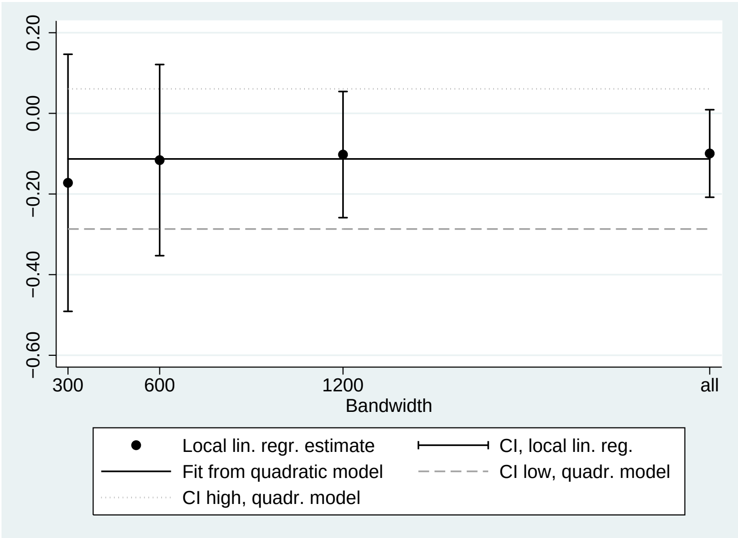


Figure 26: Local linear regression estimates for varying bandwidth and estimate from quadratic model for APGAR<=7 at 5 min

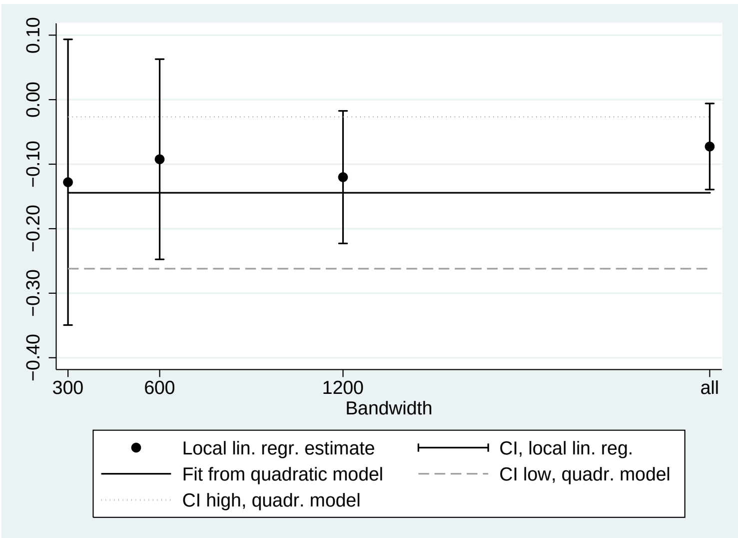


Figure 27: Local linear regression estimates for varying bandwidth and estimate from quadratic model for child morbidity

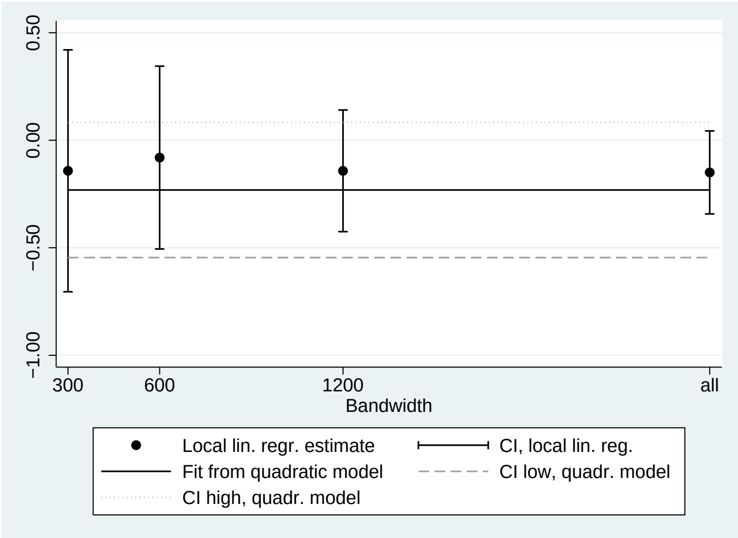


Figure 28: Local linear regression estimates for varying bandwidth and estimate from quadratic model for above average number of GP visits

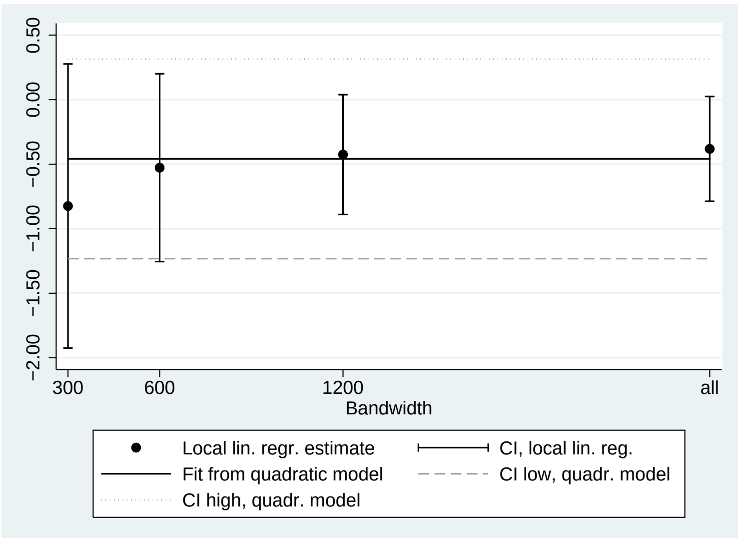


Table 1: Summary statistics, all non-breech singletons and breech pregnancies with parity>1, means and standard deviations.

	Non-breech	*Breech*
N	332809	8307
	mean/sd	mean/sd
CS	0.126	0.722
Elective CS	0.071	0.479
Emergency CS	0.045	0.193
Att. external cephalic version	0.011	0.237
APGAR<=7 at 1 min	0.022	0.037
APGAR<=7 at 5 min	0.008	0.012
Death in first year	0.001	0.004
Serious morbidity, child	0.077	0.123
Average birth hospitalization	3.131	4.567
	4.382	5.085
Birth weight	3643.797	3440.916
	510.327	539.321
Mom's age	31.058	31.412
	4.380	4.388
Mom high education	0.344	0.331
Pregnancy complications	0.250	0.336
Pre-eclampsia	0.029	0.034
Diabetes	0.019	0.020
Post-birth complications, mother	0.053	0.060
Infection	0.007	0.017
Hosp.>3 at birth	0.168	0.423
Hosp.>6 at birth	0.022	0.064
Hosp.>3 in first 3 years	0.125	0.159

Continued on the next page.

Table 1 *continued.*

	Non-breech	Breech
Hosp.>6 in first 3 years	0.069	0.088
Hosp.>3 in first 5 years	0.139	0.174
Hosp.>6 in first 5 years	0.077	0.098
Out patient>1 in first 3 years	0.209	0.244
Out patient>3 in first 3 years	0.087	0.115
Out patient>5 in first 3 years	0.048	0.068
Out patient>3 in first 5 years	0.135	0.163
Out patient>5 in first 5 years	0.081	0.103
GP visits in first 2 years	16.003	16.622
	11.273	12.171
N	234028	5766
GP visits>17 in first 2 years	0.355	0.378
Visits at specialist in first 2 years	0.621	0.658
	1.680	1.712
N	234028	5766
Specialist visits>3	0.042	0.049

Table 2: Summary statistics, breech singletons with parity>1, means and standard deviations.

	Before TBT	*After TBT*
CS	0.623	0.786
N	3267	5040
Elective CS	0.389	0.537
Emergency CS	0.159	0.215
Att. external cephalic version	0.219	0.248
APGAR<=7 at 1 min	0.039	0.036
APGAR<=7 at 5 min	0.014	0.011
Death in first year	0.003	0.004
Serious morbidity, child	0.117	0.127
Average birth hospitalisation	4.934	4.330
	5.612	4.697
Birth weight	3431.206	3447.210
	556.645	527.748
Mom's age	31.047	31.649
	4.341	4.403
Mom high school education	0.062	0.061
Pregnancy complications	0.325	0.343
Pre-eclampsia	0.036	0.033
Diabetes	0.017	0.022
Post-birth complications, mother	0.063	0.057
Infection	0.020	0.015
Hosp.>3 at birth	0.527	0.355
Hosp.>6 at birth	0.092	0.046
Hosp.>3 in first 3 years	0.185	0.142
Hosp.>6 in first 3 years	0.100	0.081
Hosp.>3 in first 5 years	0.202	0.156
Hosp.>6 in first 5 years	0.111	0.090

Continued on the next page.

Table 2 *continued.*

	Before TBT	After TBT
Out patient>1 in first 3 years	0.185	0.282
Out patient>3 in first 3 years	0.084	0.135
Out patient>5 in first 3 years	0.047	0.081
Out patient>3 in first 5 years	0.123	0.189
Out patient>5 in first 5 years	0.072	0.124
GP visits in first 2 years	16.376	16.942
	12.670	11.481
N	3267	2499
GP visits>17 in first 2 years	0.366	0.394
N	3267	2499
Visits at specialist in first 2 years	0.666	0.647
	1.816	1.566
N	3267	2499
Specialist visits>3	0.050	0.048
N	3267	2499

Table 3: ITT results for the indicator for post-TBT birth, breech singletons with parity>1, various outcomes.

Outcome measure	*Quadratic*	*Quadratic, control*
Low APGAR at one min	-0.016	-0.016
	(0.013)	(0.013)
N	8307	8307
Low APGAR at 5 min	-0.021***	-0.021***
	(0.008)	(0.008)
N	8307	8307
Serious morbidity	-0.034	-0.033
	(0.022)	(0.022)
N	8307	8307
Hosp >3 at birth	0.030	0.029
	(0.034)	(0.034)
N	8307	8307
Hosp >6 at birth	0.001	0.001
	(0.018)	(0.018)
N	8307	8307
Hosp >3 in first 3 years	0.022	0.024
	(0.025)	(0.025)
N	8307	8307
Hosp >6 in first 3 years	-0.003	-0.001
	(0.020)	(0.020)
N	8307	8307
Hosp >3 in first 5 years	0.022	0.025
	(0.026)	(0.026)
N	8307	8307
Hosp >6 in first 5 years	-0.011	-0.009
	(0.021)	(0.021)

Continued on the next page.

Table 3 *continued.*

Outcome measure	*Quadratic*	*Quadratic, control*
N	8307	8307
Out patient>3 in first 3 years	0.014	0.018
	(0.027)	(0.027)
N	8307	8307
Out patient>3 in first 5 years	-0.018	-0.017
	(0.014)	(0.014)
N	8307	8307
Out patient>5 in first 5 years	-0.036**	-0.034*
	(0.017)	(0.017)
N	8307	8307
Mom: infection	0.005	0.005
	(0.009)	(0.009)
N	8307	8307
Mom: complication post birth	0.000	0.001
	(0.016)	(0.016)
N	8307	8307
GP>17 in first 2 years	-0.051	-0.043
	(0.040)	(0.039)
N	5766	5766
Specialist>3 in first 2 years	0.018	0.020
	(0.017)	(0.017)
N	5766	5766

Notes: Table shows estimates for the effect of post-TBT birth. All cells are different regressions that additionally include quadratic polynomials and their interactions of the post-TBT dummy. Regressions are based on full sample. Column 1 does not include controls, column 2 controls for child's sex, a set of indicators for maternal age and educational group, and a set of maternal pregnancy complications.

Table 4: Local linear regression estimates for the jump in the probability of having a C-section for various bandwidths.

Estimate of jump, bw 15	0.025
	(0.237)
Estimate of jump, bw 30	0.208
	(0.162)
Estimate of jump, bw 60	0.128
	(0.122)
Estimate of jump, bw 90	0.097
	(0.096)
Estimate of jump, bw 120	0.121
	(0.081)
Estimate of jump, bw 150	0.092
	(0.074)
Estimate of jump, bw 180	0.140*
	(0.072)
Estimate of jump, bw 210	0.149**
	(0.067)
Estimate of jump, bw 240	0.130**
	(0.057)
Estimate of jump, bw 270	0.147**
	(0.059)
Estimate of jump, bw 300	0.150***
	(0.054)
Estimate of jump, bw 330	0.140***
	(0.053)
Estimate of jump, bw 360	0.143***
	(0.049)

Notes: Local linear regression is estimated with a rectangular kernel. Table presents estimates of the jump at the discontinuity for various bandwidths. Bootstrapped standard errors in parentheses (200 replications).

Table 5: First stage parametric regression: Coefficient and std. error for treatment indicator. Alternative data windows and parametric specifications.

	RoT
Treated	0.113***
	(0.043)
N	1938
F-value	7.005
	RoT, quadratic
Treated	0.172***
	(0.065)
N	1938
F-value	6.953
	Local, quadratic
Treated	0.132***
	(0.039)
N	5347
F-value	11.660
	Quadratic
Treated	0.145***
	(0.032)
N	8307
F-value	20.765
	Quadratic, control
Treated	0.145***
	(0.032)
N	8307
F-value	20.710
	Cubic
Treated	0.139***
	(0.043)
N	8307
F-value	10.477
	Cubic, control
Treated	0.135***
	(0.043)
N	8307
F-value	9.927

Notes: Local specifications include the data as graphed above (40 bins of 30 days); RoT bandwidth as suggested for the rectangular kernel by Fan and Gijbels (1996); Controls include child's sex, maternal age and educational group, and a set of maternal pregnancy complications. Robust standard errors. ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table 6: Effect C-section on short-run child health outcomes: APGAR<=7 at 1 and 5 minutes, indicator for serious perinatal morbidity

	plain	*contr*	*plain, cubic*	*contr, cubic*	*plain, small*	*contr, small*	*loclin,double ROT*	*loclin, ROT*
				APGAR<=7 at 1 min				
CS	-0.113	-0.115	-0.064	-0.069	-0.149	-0.158	-0.105	-0.101
	(0.089)	(0.089)	(0.120)	(0.123)	(0.118)	(0.121)	(0.094)	(0.145)
N	8307	8307	8307	8307	5347	5347	3917	1938
				APGAR<=7 at 5 min				
CS	-0.144**	-0.145**	-0.122	-0.125	-0.146*	-0.151*	-0.143**	-0.115
	(0.060)	(0.060)	(0.084)	(0.087)	(0.079)	(0.082)	(0.064)	(0.097)
N	8307	8307	8307	8307	5347	5347	3917	1938
				Serious perinatal morbidity				
CS	-0.231	-0.221	0.033	0.029	-0.099	-0.100	-0.133	-0.008
	(0.160)	(0.160)	(0.212)	(0.218)	(0.205)	(0.209)	(0.170)	(0.256)
N	8307	8307	8307	8307	5347	5347	3917	1938

Notes: F-values for first stage regressions: see Table 5. Robust standard errors in parentheses; controls include: child's sex, a set of indicators for maternal age and educational group, and a set of maternal pregnancy complications; Models in columns 1 and two include a second order polynomial in the forcing variable, columns 3 and 4 include a third order polynomial, columns 5 and 6 estimate the quadratic specification on the smaller 40 months sample; columns 7 and 8 present the results of a local linear regression with the rule of thumb bandwidth (and double the RoT bandwidth) suggested for the rectangular kernel by Fan and Gijbels (1996), we use the outcome APGAR<=7 at 5 minute to calculate the RoT bandwidth and use the same band width on both sides of the cut-off (447 days); ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table 7: Effect C-section on short-run child health outcomes: Child hospitalizations.

	plain	*contr*	*plain, cubic*	*contr, cubic*	*plain, small*	*contr, small*	*loclin,double ROT*	*loclin, ROT*
				hospitalizations >3 days at birth				
CS	0.204	0.195	0.035	0.013	-0.046	-0.067	0.071	-0.149
	(0.219)	(0.220)	(0.324)	(0.336)	(0.324)	(0.333)	(0.254)	(0.431)
N	8307	8307	8307	8307	5347	5347	3917	1938
				hospitalizations >6 days at birth				
CS	0.004	0.003	-0.048	-0.067	-0.128	-0.147	-0.031	-0.069
	(0.121)	(0.121)	(0.172)	(0.178)	(0.173)	(0.177)	(0.133)	(0.211)
N	8307	8307	8307	8307	5347	5347	3917	1938
				hospitalizations >3 days in first 3 years				
CS	0.149	0.169	0.229	0.227	0.268	0.282	0.191	0.082
	(0.175)	(0.176)	(0.252)	(0.257)	(0.245)	(0.248)	(0.196)	(0.302)
N	8307	8307	8307	8307	5347	5347	3917	1938
				hospitalizations >6 days in first 3 years				
CS	-0.077	-0.062	-0.007	-0.010	0.005	0.013	0.049	-0.127
	(0.144)	(0.144)	(0.200)	(0.205)	(0.190)	(0.192)	(0.156)	(0.251)
N	8307	8307	8307	8307	5347	5347	3917	1938

Notes: see Notes for Table 6. Robust std.errors; ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table 8: Effect C-section on short-run child health outcomes: Child outpatient contacts with hospitals.

	plain	*contr*	*plain, cubic*	*contr, cubic*	*plain, small*	*contr, small*	*loclin,double ROT*	*loclin, ROT*
			Outpaptient contacts >3 days in first 3 years					
CS	-0.111	-0.105	0.185	0.184	0.113	0.118	-0.018	0.106
	(0.131)	(0.132)	(0.177)	(0.182)	(0.168)	(0.172)	(0.138)	(0.209)
N	8307	8307	8307	8307	5347	5347	3917	1938
			Outpaptient contacts >5 days in first 3 years					
CS	-0.120	-0.124	0.099	0.092	0.035	0.030	-0.100	0.173
	(0.102)	(0.103)	(0.131)	(0.135)	(0.127)	(0.130)	(0.109)	(0.164)
N	8307	8307	8307	8307	5347	5347	3917	1938
			Outpaptient contacts >3 days in first 5 years					
CS	-0.226	-0.213	0.000	0.003	-0.115	-0.115	-0.097	-0.239
	(0.164)	(0.163)	(0.208)	(0.213)	(0.205)	(0.209)	(0.169)	(0.273)
N	8307	8307	8307	8307	5347	5347	3917	1938
			Outpaptient contacts >5 days in first 5 years					
CS	-0.244*	-0.241*	0.014	0.010	-0.058	-0.065	-0.188	-0.031
	(0.135)	(0.136)	(0.159)	(0.164)	(0.158)	(0.162)	(0.141)	(0.195)
N	8307	8307	8307	8307	5347	5347	3917	1938

Notes: see Notes for Table 6. Robust std.errors; ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

Table 9: Effect C-section on short-run child health outcomes: Child GP visits and contacts with a specialised physician.

	plain	*contr*	*plain, cubic*	*contr, cubic*	*plain, small*	*contr, small*	*loclin,double ROT*	*loclin, ROT*
				GP contacts >17 in first 2 years				
CS	-0.459	-0.413	-0.700*	-0.717*	-0.484	-0.449	-0.350	-0.453
	(0.394)	(0.391)	(0.382)	(0.390)	(0.389)	(0.388)	(0.271)	(0.432)
N	5766	5766	5766	5766	5171	5171	3917	1938
				Specialist contacts >3 in first 2 years				
CS	0.161	0.186	0.058	0.068	0.144	0.174	0.100	0.148
	(0.162)	(0.167)	(0.129)	(0.133)	(0.156)	(0.162)	(0.113)	(0.173)
N	5766	5766	5766	5766	5171	5171	3917	1938

Notes: see Notes for Table 6. Robust std.errors; ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

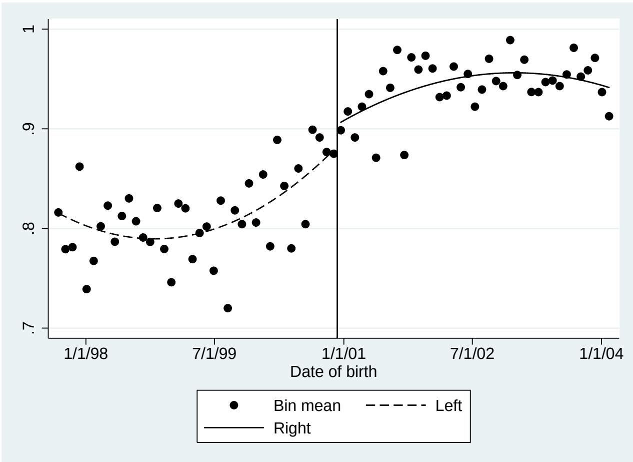
Table 10: Effect C-section on short-run mother health outcomes: post-birth infections and complications.

	plain	*contr*	*plain, cubic*	*contr, cubic*	*plain, small*	*contr, small*	*loclin,double ROT*	*loclin, ROT*
				Mom: post-birth infections				
CS	0.031	0.029	-0.012	-0.022	-0.030	-0.045	0.016	-0.085
	(0.059)	(0.059)	(0.085)	(0.088)	(0.080)	(0.083)	(0.064)	(0.110)
N	8307	8307	8307	8307	5347	5347	3917	1938
				Mom: post-birth complications				
CS	0.003	0.003	0.006	0.003	-0.040	-0.051	-0.002	0.040
	(0.110)	(0.111)	(0.155)	(0.159)	(0.152)	(0.155)	(0.123)	(0.191)
N	8307	8307	8307	8307	5347	5347	3917	1938

Notes: see Notes for Table 6. Robust std.errors; ***significant at the 1 pct level, **significant at the 5 pct level *significant at the 10 pct level

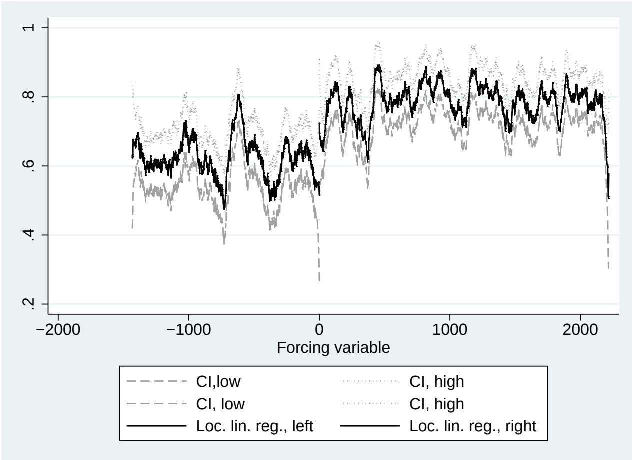
A Appendix

Figure A.1: C-section rate for breech pregnancies, parity==1



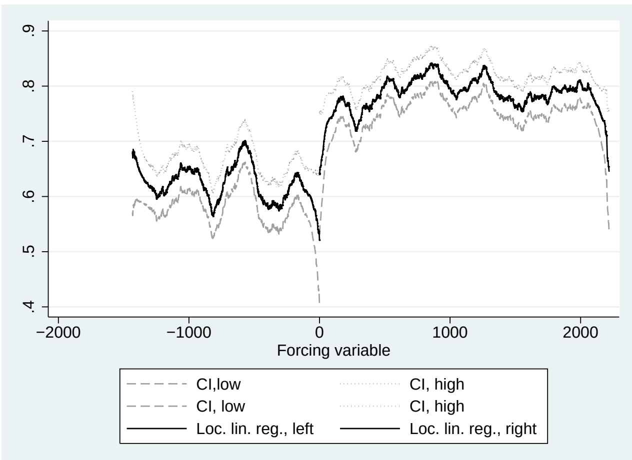
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.2: Discontinuity in probability of having a C-section



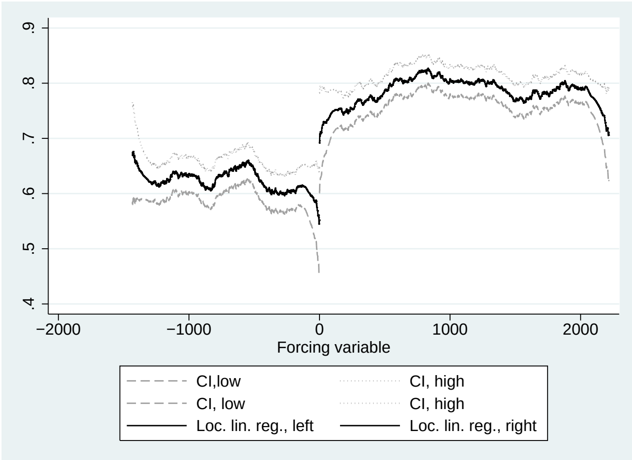
Notes: Local linear regression with confidence bands, using a rectangular kernel and bandwidth 30 days.

Figure A.3: Discontinuity in probability of having a C-section



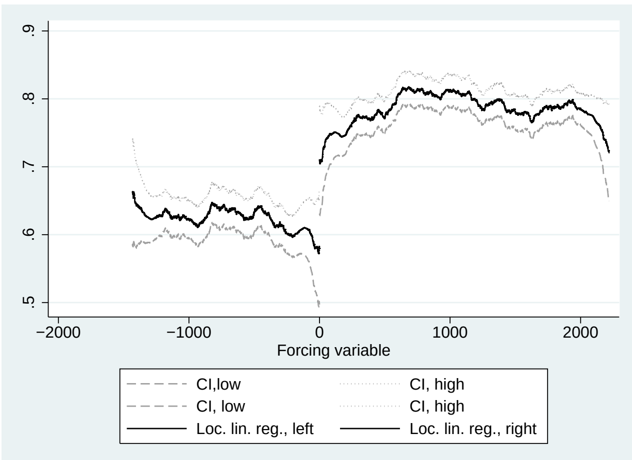
Notes: Local linear regression with confidence bands, using a rectangular kernel and bandwidth 120 days.

Figure A.4: Discontinuity in probability of having a C-section



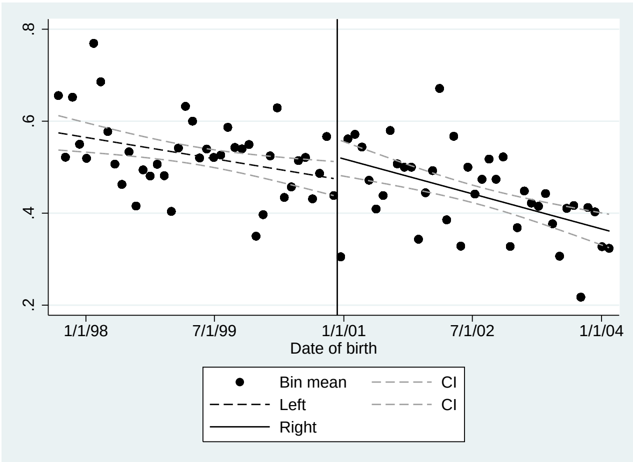
Notes: Local linear regression with confidence bands, using a rectangular kernel and bandwidth 180 days.

Figure A.5: Discontinuity in probability of having a C-section



Notes: Local linear regression with confidence bands, using a rectangular kernel and bandwidth 240 days.

Figure A.6: Percentage of children with more than 3 days of hospitalization at birth



Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.7: Percentage of children with more than 3 days of hospitalization in first 3 years

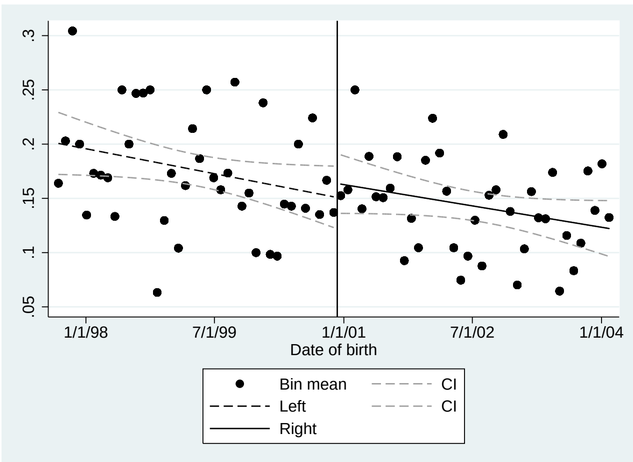
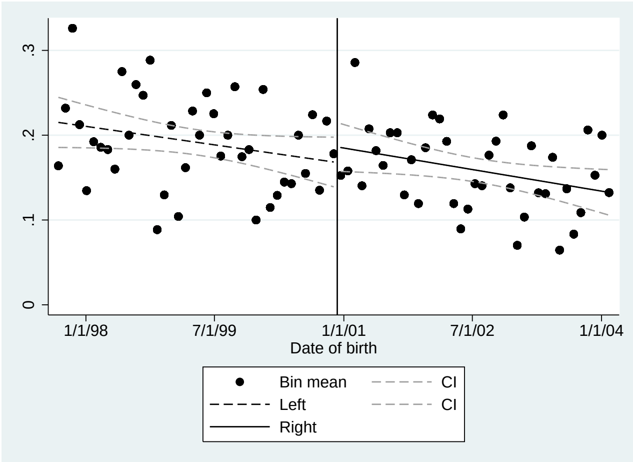
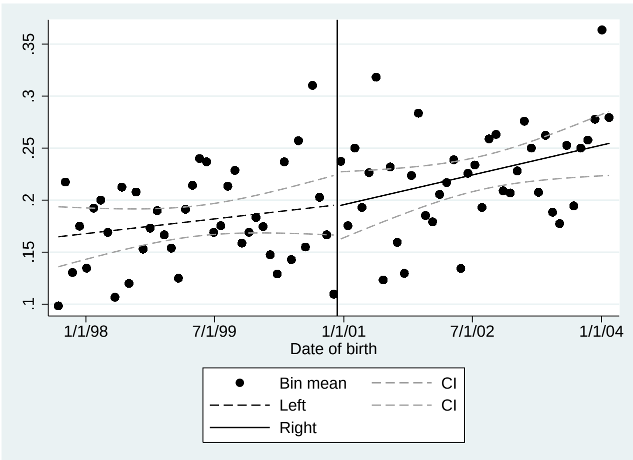


Figure A.8: Percentage of children with more than 3 days of hospitalization in first 5 years



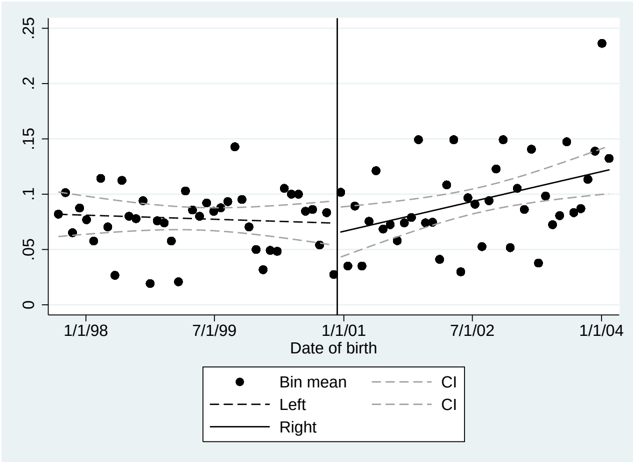
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.9: Percentage of children with more than 1 outpatient visits in first 3 years



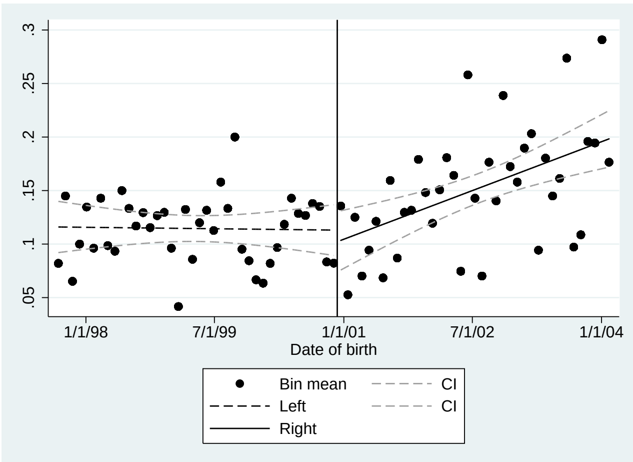
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.10: Percentage of children with more than 3 outpatient visits in first 3 years



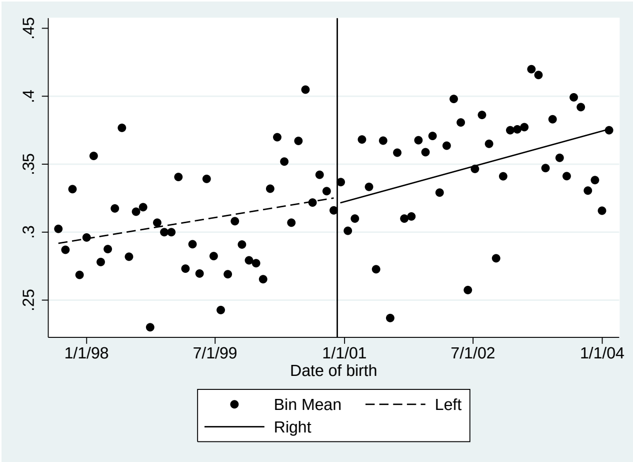
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.11: Percentage of children with more than 3 outpatient visits in first 5 years



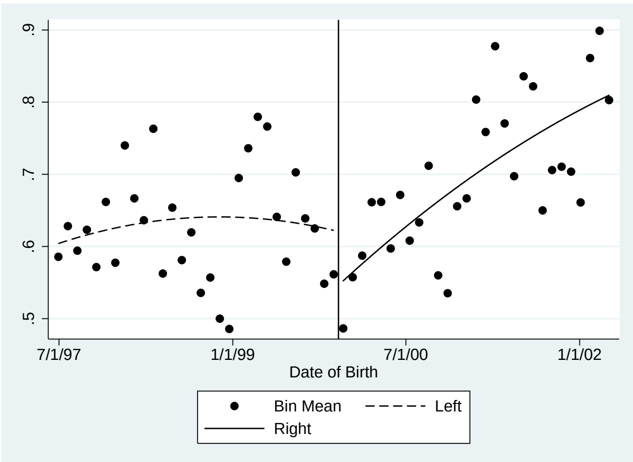
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.12: C-section rate for pregnancies with pre-eclampsia



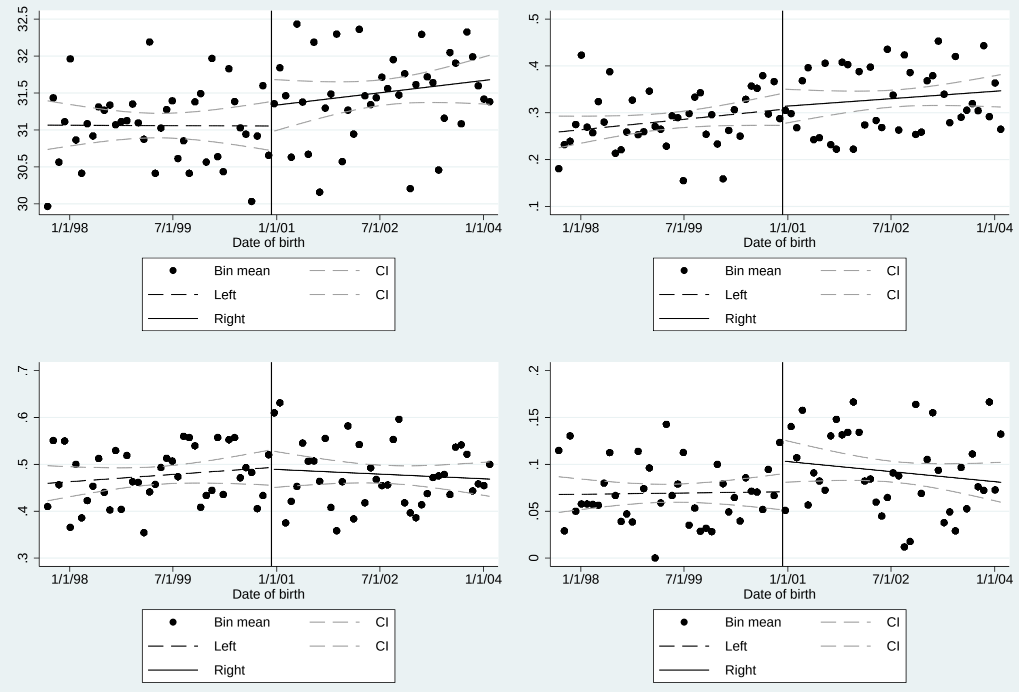
Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.13: C-section rate for breech pregnancies, parity>1: Placebo cut-off: 1999



Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.14: Test for jump in observable characteristics



Graphs for mom's age, Share of high edu. moms, Share of male children, Share of immigrant moms

Notes: Width of bins: 30 days, 40 bins on each side of the cut-off.

Figure A.15: IV estimates and CI for APGAR<=7 at 1 min for different bandwidths

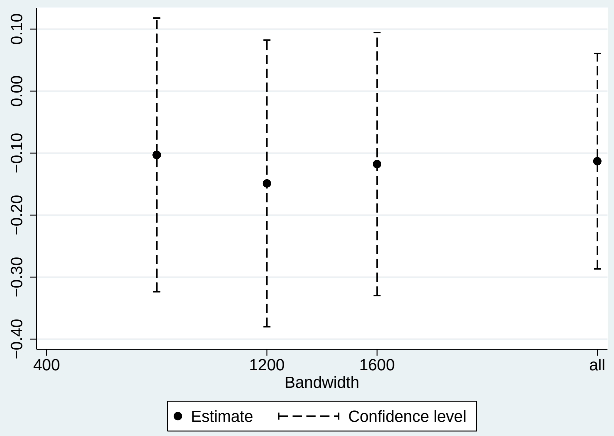


Figure A.16: IV estimates and CI for APGAR<=7 at 5 min for different bandwidths

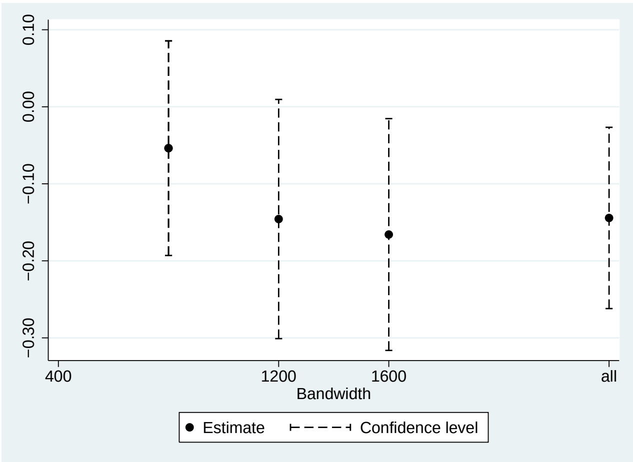


Figure A.17: IV estimates and CI for Morbidity for different bandwidths

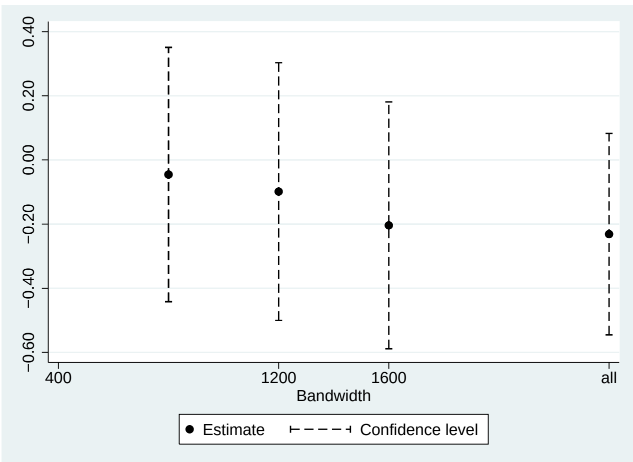


Figure A.18: IV estimates and CI for GP visits over mean number for different bandwidths

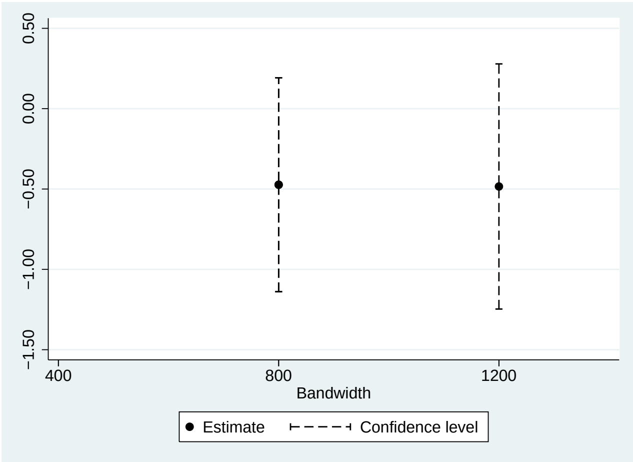


Figure A.19: IV estimates and CI for hospitalizations more than 3 days after birth for different bandwidths

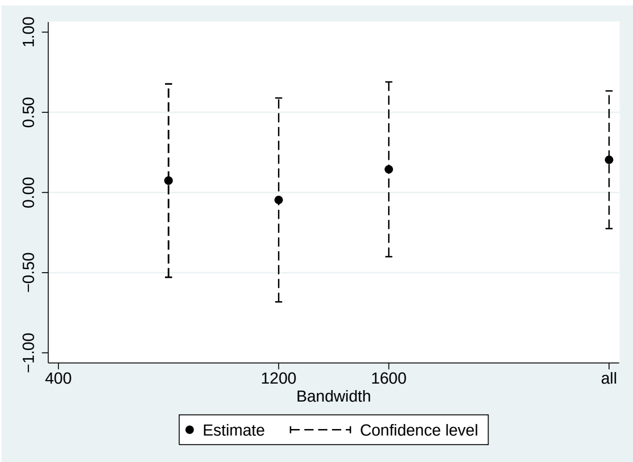


Figure A.20: IV estimates and CI for hospitalizations more than 3 days in first 3 years for different bandwidths

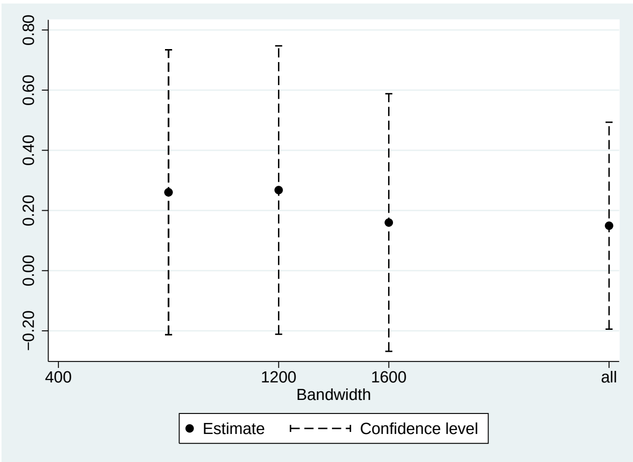


Table A.1: Choice of bandwidth in graphs for first stage and outcomes

	P-values of F-tests
	First Stage
bintest 40	.0624344
N	5209
regtest 60	.166961
N	5209
bintest 80	.0792647
N	5277
regtest 30	.3998399
N	5277
	APGAR<=7 at 5 min
bintest 40	.3687139
N	5209
regtest 60	.3132326
N	5209
bintest 80	.4168955
N	5277
regtest 30	.6464323
N	5277
	Hospitalization >3 days in first 3 years
bintest 40	.61213
N	5209
regtest 60	.3595614
N	5209
bintest 80	.1373322
N	5277
regtest 30	.2323503
N	5277
	GP visits >17 in first 2 years
bintest 40	.4365686
N	5168
regtest 60	.2776461
N	5168
bintest 80	.1630554
N	5168
regtest 30	.1017875
N	5168

Notes: Consult section 4 for a further description of both tests. Models include 20 bins with bandwidth 60 on each side of the cut-off and 40 bins with bandwidth 30, respectively. The p-values for other outcomes are similar and available on request.

Table A.2: Rule of Thumb bandwidth in a local linear regression for treatment and outcomes

	RoT bandwidth
	.
RoT CS	
RoT left	341.846
RoT right	411.131
	.
RoT APGAR<=7 at 5 min	
RoT left	321.247
RoT right	447.063
	.
RoT Morbidity	
RoT left	988.855
RoT right	603.946
	.
RoT GP visits indicator	
RoT left	237.148
RoT right	231.559

Notes: Consult section 5 for further description of RoT bandwidth. The table shows the RoT bandwidths estimated separately for both sides of the cut-off. In the main analysis we use the RoT bandwidth for the APGAR score at five minutes for all outcomes for convenience.

DEPARTMENT OF ECONOMICS AND BUSINESS

AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
www.econ.au.dk

PhD Theses since 1 July 2011

2011-4 Anders Bredahl Kock: Forecasting and Oracle Efficient Econometrics

2011-5 Christian Bach: The Game of Risk

2011-6 Stefan Holst Bache: Quantile Regression: Three Econometric Studies

2011:12 Bisheng Du: Essays on Advance Demand Information, Prioritization and Real Options in Inventory Management

2011:13 Christian Gormsen Schmidt: Exploring the Barriers to Globalization

2011:16 Dewi Fitriasari: Analyses of Social and Environmental Reporting as a Practice of Accountability to Stakeholders

2011:22 Sanne Hiller: Essays on International Trade and Migration: Firm Behavior, Networks and Barriers to Trade

2012-1 Johannes Tang Kristensen: From Determinants of Low Birthweight to Factor-Based Macroeconomic Forecasting

2012-2 Karina Hjortshøj Kjeldsen: Routing and Scheduling in Liner Shipping

2012-3 Soheil Abginehchi: Essays on Inventory Control in Presence of Multiple Sourcing

2012-4 Zhenjiang Qin: Essays on Heterogeneous Beliefs, Public Information, and Asset Pricing

2012-5 Lasse Frisgaard Gunnersen: Income Redistribution Policies

2012-6 Miriam Wüst: Essays on early investments in child health

ISBN 8790117883