Universitat Pompeu Fabra
Departament d'Economia i Empresa

DOCTORAT EN ECONOMIA, FINANCES I EMPRESA

Essays on Child Health and Family Economics

Tesi Doctoral de:
Ana Costa-Ramón

Directors de la tesi:
Libertad González
Guillem López

TESI DOCTORAL UPF / 2020

ACKNOWLEDGEMENTS

I am extremely grateful to Libertad for her guidance and support throughout these years. Libertad's work has inspired and shaped my research. Her advice and encouragement have meant a lot to me. I am also indebted to Guillem for trusting and guiding me from early on.

I am also grateful to Alessandro Tarozzi, Ana Tur, Christian Fons-Rosen, Vicente Ortún, and everyone in the UPF LPD group for their time and valuable comments. I would also like to thank Laura Agustí and Marta Araque for always helping and supporting us in the best possible way.

A special thank you to Ana Rodríguez, for being the best friend and co-author that one could ever have, for sharing everything with me. My best memories of these years are with you, learning by your side has been a true pleasure. Many thanks to my UPF friends: Christian, Christoph, Dani, Korie, Maitane, Matt, Óscar, Patricia, and Xulia. To Andrea, Irene, and Tanit, for always being there.

I am deeply indebted to my parents, Rafael and María, and my sister, Cristina, because they made this possible and always believed it was. And last, to Miquel, because he suffered and celebrated every step of this process with me, and I am very lucky to share the past, the present, and my future with him.

ANA COSTA-RAMÓN
Barcelona
Juny 2020

Abstract

This dissertation consists of three essays that investigate the effect of children's health shocks on children's and parent's well-being. In the first chapter, we investigate the effect of C-sections on newborn health. We use variation in the probability of unplanned C-section by the time of day as an instrument for type of delivery and find a small negative impact on neonatal health. In the second chapter, we use a similar methodology to analyze the long-term effects of C-sections. We find that unplanned C-sections increase the risk of asthma, but do not affect other immune-mediated disorders previously associated with C-sections. In the last chapter, I study the impact of a child's adverse health event on parental labor market outcomes. I do this by comparing parents across families in similar parental and child age cohorts whose children experienced a health shock at different ages. I show that parental earnings suffer a substantial and persistent decline after the event. I also find an impact on parents' mental well-being.

Resum

Aquesta tesi està formada per tres assajos que investiguen l'efecte dels xocs en la salut dels infants, en el seu benestar i el de la seva família. Al primer capítol, investiguem l'efecte de néixer per cesària en la salut neonatal. Utilitzem variació en la probabilitat de cesària no planificada segons l'hora del dia com a instrument pel tipus de part i trobem un efecte negatiu, petit, en la salut neonatal. Al segon capítol, utilitzem una metodologia similar per analitzar l'efecte de néixer per cesària a llarg termini. Trobem que néixer per cesària no planificada augmenta el risc de patir asma, però no afecta altres malalties immunològiques que prèviament s'havien trobat associades amb la cesària. A l'últim capítol estudio l'impacte que té que l'infant pateixi un xoc en la seva salut, al mercat laboral de les mares i pares. La meva estratègia d'identificació es basa a comparar progenitors amb la mateixa edat, amb fills de la mateixa edat, però que pateixen el xoc en diferents moments. Els ingressos de les mares i pares pateixen una caiguda substancial i persistent després de l'episodi. També trobo que aquest esdeveniment afecta la salut mental de les mares i pares.

RESUMEN

Esta tesis está formada por tres ensayos que investigan el efecto de *shocks* en la salud de los niños y niñas, en su bienestar y el de su familia. En el primer capítulo, investigamos el efecto de nacer por cesárea en la salud neonatal. Utilizamos variación en la probabilidad de cesárea no planificada según la hora del día como instrumento para el tipo de parto, y encontramos un efecto negativo, pequeño, en la salud neonatal. En el segundo capítulo, utilizamos una metodología similar para analizar el efecto de nacer por cesárea a largo plazo. Encontramos que nacer por cesárea no planificada aumenta el riesgo de sufrir asma, pero no afecta otras enfermedades inmunológicas que previamente se asociaban con nacer por cesárea. En el último capítulo estudio el impacto de sufrir una hospitalización severa durante la infancia, en el mercado laboral de las madres y padres. Mi estrategia de identificación se basa en comparar progenitores con la misma edad, con hijos de la misma edad, pero que sufren el evento en diferentes momentos. Los ingresos de las madres y padres sufren una caída sustancial y persistente después del episodio. También encuentro que esto afecta a la salud mental de las madres y padres.

Preface

This dissertation consists of three chapters that investigate the effect of children's health shocks on children's and family's well-being. To do so, I exploit rich administrative data sources and apply cutting-edge econometric techniques that allow me to contribute to the previous literature by providing credible causal estimates of the impact of health shocks during childbirth, and from childhood to teenage years.

In the first two chapters, I study the causal impact of being born by cesarean section on neonatal and infant health. Cesarean sections have been associated in the literature with poorer newborn and infant health. Most studies suffer, however, from potential omitted variable bias, as they are based on simple comparisons of mothers who give birth vaginally with those who give birth by cesarean section. These two papers overcome this limitation by exploiting different sources of variation in the probability of C-section, which are unrelated to maternal and child characteristics.

In the first chapter, co-authored with Ana Rodríguez-González, Miquel Serra-Burriel, and Carlos Campillo, we investigate the impact of C-sections on newborn health. Using a sample of hospitals in Spain, we first show that the rate of unplanned C-sections is higher during the early hours of the night compared to the rest of the day. We use this variation as an instrument for the type of birth. We find a small negative impact on neonatal health, as measured by Apgar Scores, but the effect is not severe enough to translate into more extreme outcomes.

In the second chapter, joint work with Mika Kortelainen, Ana Rodríguez-González, and Lauri Sääksvouri, we continue with the same line of research and study the long-term effects of C-sections. Using Finnish administrative data, we document that physicians perform more unplanned C-sections during their regular working hours on days that precede a weekend or a public holiday and use this exogenous variation as an instrument for C-sections. We supplement our instrumental variables results with a differences-in-differences estimation that exploits variation in birth mode within sibling pairs and across families. We find that avoidable unplanned C-sections increase the risk of asthma, but do not affect other immune-mediated disorders previously associated with C-sections.

In the last chapter of my thesis, I study the causal impact of children's severe health shocks on parental labor market outcomes and mental well-being. Although economists have long been interested in understanding the relationship between income and health, we know relatively little about the potential spillover effects of health shocks on other family members. The illness of a child is a stressful event that can have major implications for the well-being of the whole household. Families can incur substantial costs when deciding how to best cope with these health shocks and their associated long-term burden. In this paper, I contribute to filling this gap by providing new credible causal evidence on the impact of children's adverse health events on parental labor outcomes. To do so, I leverage long panels of high-quality Finnish administrative data combined with a research design that allows me to overcome previous limitations. In particular, identification comes from comparisons among parents of the same age cohorts, with children of the same age, whose children experienced the health shock at different ages. This enables me to abstract from differences across treated and untreated families. My results show that parental earnings suffer a substantial and persistent decline following a child's adverse health event. I also find that these shocks impact parents' mental well-being.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

IT'S ABOUT TIME: CESAREAN SECTIONS AND NEONATAL HEALTH

Joint with Ana Rodríguez-González (UPF), Miquel Serra-Burriel (CRES-UPF) and Carlos Campillo-Artero (Servei de Salut de les Illes Balears)

Published as: Costa-Ramón, A. M., Rodríguez-González, A., Serra-Burriel, M., and Campillo-Artero, C. (2018). It's about time: Cesarean sections and neonatal health. *Journal of Health Economics*, 59:46 – 59.
DOI: https://doi.org/10.1016/j.jhealeco.2018.03.004
www.sciencedirect.com/science/article/abs/pii/S0167629617307609

1.1 INTRODUCTION

Recent years have seen increasing concern over the rise in cesarean section births. Among OECD countries in 2013, on average more than 1 out of 4 births involved a c-section, compared to 1 out of 5 in 2000 (OECD, 2013). This rise has been largely debated because c-sections are associated with greater complications and higher maternal and infant mortality and morbidity compared to vaginal births. However, the available studies may suffer from omitted variable bias, as mothers who give birth by c-sections may be different from those who have vaginal births in terms of characteristics that can affect the health outcomes of the child and the mother after birth. Along these lines, the WHO has recently pointed out the need for more research in order to better understand the health effects of cesarean sections on immediate and future outcomes, remarking that "the effects of cesarean section rates on other outcomes, such as maternal and neonatal morbidity, pediatric outcomes and psychological or social well-being, are still unclear" (WHO, 2015).

This paper aims to help fill this research gap by providing new evidence of a causal link between unplanned cesarean sections and newborn health outcomes. Understanding the impact of c-sections on neonatal health is of relevance, as fetal and neonatal outcomes have been shown to be determinants not only of future health, but also of other later life outcomes, such as test scores, educational attainment, and income (Almond and Currie, 2011). In particular, we look at the impact of c-sections on Apgar scores, a widely used measure of newborn well-being. Apgar scores have been found to be predictive of health, cognitive ability, and behavioral problems of children at age three (Almond et al., 2005), of reading and math test scores in grades 3-8 (Figlio et al., 2014), and of school attainment and social assistance receipt after age 18 (Oreopoulos et al., 2008). We also analyze the effect of c-sections on other indicators of newborn wellbeing, such as needing reanimation or being admitted to the intensive care unit (ICU).

In order to show the existence of a causal relationship between unscheduled c-sections and health, we use exogenous variation in the probability of having a c-section at different times of day. Indeed, although nature distributes births and associated problems uniformly, some studies have demonstrated that time-dependent variables related to physicians' demand for leisure are significant predictors of unplanned c-sections (Brown, 1996). Using a sample of birth registries in public hospitals in Spain, we first document that, in this context, unplanned c-sections are more likely to be performed in the early hours of the night (from 11 pm to 4 am). We discuss how the structure of medical shifts and the higher opportunity cost in terms of time that vaginal deliveries imply might explain physicians' incentives to perform more c-sections during this time of day. We then show that mothers giving birth at different times of day are observationally similar, also in terms of pregnancy and labor characteristics that might predict a medically-indicated c-section. The results thus suggest that the excess number of c-sections observed at the early night are due to non-medical reasons. We consequently adopt an instrumental variable approach, using time of birth as an instrument for the mode of delivery. In other words, we estimate the local average treatment effect of c-sections on neonatal health for mothers whose mode of delivery is affected by time of birth. This allows us to interpret our estimates as causal and to focus on avoidable c-sections, as medically-indicated cesareans will be performed independently of the time of birth. Our results suggest that these non-medically indicated c-sections lead to a significant worsening of Apgar scores of approximately one standard deviation, but we do not find effects on more extreme outcomes such as needing reanimation, being admitted to the ICU or on neonatal death.

In order for our instrument to be valid, it must satisfy two conditions: first, that there is no selection of mothers with different characteristics giving birth at different times of day and, second, that giving birth during the early hours of the night only affects infant health through the increased probability of having a c-section. The comparison of maternal and pregnancy characteristics across times of day provides reassuring evidence regarding the first assumption. In order to support the validity of the exclusion restriction and, in particular, to show that variation in quality of care across time is not driving our results, we perform a robustness check restricting the analysis to births that take place during the night. Moreover, section 5 includes further supplementary tests that support our interpretation of the findings.

This paper contributes to two different strands of the literature. First, we contribute to studies on the effects of c-sections on newborn health outcomes. A large number of papers have documented a robust association between c-sections and respiratory morbidity, both at birth (Zanardo et al., 2004; Hansen et al., 2008) and in the longer-term in the form of asthma (Davidson et al., 2010; Sevelsted et al., 2015).

To the best of our knowledge, the only paper that endeavors to identify the causal impact of cesareans on later infant health is Jachetta (2015)[1]. The author uses variation in medical malpractice premia at the Metropolitan Statistical Area (MSA) level in the US as an instrument for the rate of risk-adjusted cesarean sections and finds that higher rates lead to an increase in the rate of total hospitalizations and of hospitalizations that present asthma. Although the author identifies several potential threats to the validity of the instrument, the paper is a first step towards providing evidence of the causal link between c-sections and health outcomes. We advance the existing knowledge by using a new instrument that allows us to credibly isolate the causal impact of non-medically indicated c-sections on newborn health. In particular, our setting allows us to focus on mothers that give birth in the same hospital and have similar observable characteristics, differing only in the time of delivery. Moreover, because we measure the impact on health at birth, we are able to establish a direct connection between c-sections and health outcomes.

Second, our work is also related to the literature that documents or uses time variation in the probability of having a c-section. Brown (1996) was one of the first to show that the probability of unplanned c-sections is non-uniformly distributed across time. Using data from military hospitals in the US, the author finds that cesarean sections

[1] Recent work by Jensen and Wüst (2015) and Mühlrad (2017) examines the impact of medically necessary c-sections on health for a particular group of at-risk babies: those in breech position at term. Their findings suggest positive short and long-run effects of medically indicated cesareans for this group.

were less likely to occur during the weekend and more likely from 6 pm to 12 am. He interprets these results as evidence that non-clinical variables, in particular physicians' demand for leisure, also play a role in doctors' decision-making. In our setting, we find that the probability of unplanned c-sections is higher during the early hours of the night. It is during this time that doctors appear to have a higher incentive to perform a c-section when facing ambiguous cases, as the opportunity cost in terms of time for a vaginal delivery is higher.

There is one paper that uses time variation in the probability of having a c-section to study maternal outcomes. Halla et al. (2016) use administrative data from Austria to show that the probability of a c-section birth is lower on weekends and public holidays. They use this as an instrument for mode of delivery, and find that c-sections reduce subsequent fertility and that this translates into an increase in maternal labor supply over a period of about six years. Our paper also makes use of time variation but our data allow us to use finer variation and rule out potential exogeneity problems: we study mothers in the same hospital, on the same day, but giving birth at different times. Moreover, we are also able to precisely identify and restrict our sample to non-scheduled c-sections.

The structure of the rest of the paper is as follows. In the next section we provide background information on the choice of mode of delivery, on the institutional setting and physicians' shifts, and on why we would expect to find an adverse effect of c-sections on health outcomes. The third section introduces the data, describes the variation in the c-section rate across a 24-hour cycle and presents the empirical strategy. In section 1.4 we show and discuss our results. Section 1.5 presents some robustness checks and supplementary analysis and, finally, section 1.6 concludes.

1.2 BACKGROUND

1.2.1 *Choice of the mode of delivery*

Cesarean sections can be performed for several reasons and at different lengths of pregnancy. First, c-sections can be scheduled in advance – also known as planned c-sections – when there are medical indications that make a vaginal delivery inadvisable. Examples of such indications include multiple pregnancies with non-cephalic presentation of the first twin or placenta previa (NICE, 2016). In principle, c-sections can also be scheduled if they are demand-determined; that is, if the mother requests to deliver via a c-section. However, in the context of public hospitals in Spain, these elective c-sections are very uncommon and are not, in fact, included in the portfolio of services offered by the public system (Marcos, 2008). In any case, we exclude scheduled c-sections from our sample as these women are likely to be different from those delivering vaginally.

If there is no scheduled c-section, an attempt of vaginal delivery begins with the onset of labor or medical induction. If an immediate threat to the life of the woman or fetus emerges, a c-section should be performed as quickly as possible (NICE, 2011). However, some indications such as dystocia (failure to progress or cephalopelvic disproportion) have a more imprecise diagnosis which leaves the door open to a more discretionary interpretation and present large variability among clinicians (Fraser et al., 1987; Barber et al., 2011). Therefore, in some cases, whether or not a c-section is needed is not obvious, and the choice between a vaginal delivery or a c-section will depend on the subjective assessment of the doctor. Unfortunately, our data does not contain the specific indication registered by the medical team to justify the c-section. However, given that emergencies should be uniformly distributed across time, we expect any observed time variation in the c-section rate to be due to indications falling in this gray area.

As Shurtz (2013) points out, a c-section is a common procedure known to be sensitive to physician incentives. Several papers have found, for example, that financial fees can influence doctors' behavior (Grant, 2009). When fees are higher for a c-section than for a vaginal delivery, physicians have a greater incentive to perform a c-section. Other studies suggest that physicians perform more c-sections as a defensive strategy reflecting a fear of malpractice lawsuits (Baicker et al., 2006; Currie and MacLeod, 2008; Jachetta, 2015). Finally, physicians have more incentives to perform c-sections when the opportunity cost of time is higher, as vaginal deliveries take longer than c-sections and thus the latter can be seen as a time-saving device (Lefèvre, 2014). We focus here on this last type of incentive given that, by performing our analysis within hospital and exploiting variation across time of day, we abstract from variations in malpractice premia and financial fees.

In particular, the average duration of vaginal deliveries among first-time mothers is around 11 hours (NICE, 2014). The first stage of established labor[2] usually lasts about 8 hours and is rarely longer than 18 hours. After that, birth is expected to take place within 3 hours of the start of the active second stage[3]. In contrast, a c-section takes much shorter; in general the average duration of this procedure is between 30 and 75 minutes (NICE, 2014). The baby is usually delivered in the first 5-15 minutes, with the remaining time being used for closing the incision (APA, 2017). Moreover, complications during this procedure are very uncommon. According to NICE (2011), c-sections increase the risk of hysterectomy (14 more per 100,000) and of cardiac arrest (15 more per 10,000). Therefore, given the low risk in terms of complications and the expected time gain, doctors may have larger incentives to perform a cesarean section when the opportunity cost of time is higher.

[2]Mothers are considered to be in the first stage of established labor when the cervix has dilated to about 4 cm (NICE, 2014).

[3]The mother is considered to be in active second stage of labor when either the baby is visible, or the full dilatation of the cervix has been accomplished and one of the following conditions is satisfied: either the mother has expulsive contractions or there is active maternal effort.

1.2.2 *Mechanisms: the impact of c-sections on newborn health*

Cesarean sections have been associated with several adverse health outcomes for newborns. Hyde et al. (2012) provide an extensive review of such findings, concluding that although further research is needed, the available evidence suggests that "normal vaginal delivery is an important programming event with life-long health consequences." More specifically, the absence or modification of a vaginal delivery has been linked to several health alterations, which they classify as either short- or long-term. In what follows we summarize some of these findings, in particular those that are more relevant to understand how c-sections might affect our outcome variables. Before doing so, however, it should be noted that any negative health effect of c-sections is outweighed by its benefits when there is a clear medical necessity. For instance, in the case of breech babies, Jensen and Wüst (2015) find that c-sections decrease the probability of having low Apgar scores and the number of doctor visits in the first year of life. More generally, cesareans save lives when severe complications arise during birth.

The adverse short-term outcomes with which c-sections have been associated include the increased risk of impaired lung functioning and altered behavioral responses to stress. With regard to the former, one of the most common causes of respiratory distress among newborns is transient tachypnea or the presence of retained lung fluid. While in the amniotic sac, a baby's lungs are filled with amniotic fluid, but during labor the baby releases chemicals which, together with the pressure of the birth canal on the baby's chest, help expel the amniotic fluid from their lungs. This process does not occur when babies are born by cesarean section, such that the presence of fluid in their lungs after birth is more common. Moreover, catecholamines, one of the chemicals released by the fetus during labor, are also correlated with muscle tone and excitability. Otamiri et al. (1991) find that babies born by cesarean section responded worse to neurological tests a few days after birth. In our setting, we can proxy the impact of c-sections on these outcomes by looking at Apgar scores at minute 1 and 5 after birth, which capture, among other aspects, respiration, reflexes and muscle tone. Severe effects, in particular serious respiratory morbidity, could also be reflected in increased need for assisted ventilation or ICU admission (Grivell and Dodd, 2011).

In the longer-term, cesarean births have also been associated with a higher risk of asthma (Sevelsted et al., 2015). While one possible mechanism is change in infant microbiome as a result of not passing through the birth canal, Hyde et al. (2012) also highlight that altered lung functioning at birth may lead to the development of future respiratory problems. Finally, there is evidence that the reduction in excitability among cesarean newborns may be a sympton of further alterations in the programming of the central nervous system, as affected by the catecholamine surge at birth (Boksa and Zhang, 2008). These findings generally suggest that any health worsening at birth we detect may have long-lasting consequences.

1.2.3 *Institutional setting*

1.2.3.1 Childbirth in Spanish public hospitals

In Spain, maternity care coverage is universal under the provision of the Spanish National Health Service. Antenatal and postnatal care for women are mainly provided at local health centers by midwives, while deliveries are supervised in hospitals by teams of both midwives and obstetricians. Expectant women do not have a pre-assigned doctor or midwife for the delivery. Rather, they are assigned to the professional available at the time of admission to the hospital. During labor, women are assisted by midwives who monitor the baby, check how labor is progressing, and call a doctor if they notice any issues. If no complications arise, midwives might manage the whole delivery. However, the obstetrician is in charge of any instrumented assistance and makes decisions regarding the mode of delivery.

Women may opt for private care, but most deliveries – 8 out of 10 births – take place under the public health system (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2015). Pregnant women are in general assigned to give birth at the hospital that is closest to their residence. In big cities where there are several public hospitals, mothers can request a change in the assigned hospital through an administrative procedure. However, hospitals in our sample are located in medium-size towns in which there are no other public hospitals.

In the year 2014, the c-section rate in the public health system was 22.1%, lower than the 25.4% rate of the whole sector, combining both public and private hospitals (*ibid.*). It is important to note that within the public system, obstetricians' wages are independent of the method of delivery used or the number of c-sections performed.

1.2.3.2 Physicians' shifts

In our setting, the typical work shift for a doctor is from 8 am to 3 pm; night shifts are covered by doctors that are on duty and must stay in the hospital for 24 hours (from 8 am to 8 am next morning). All doctors younger than 55 are required by law to work these longer shifts (Ministerio de Trabajo y Asuntos Sociales, 1997). When doctors are on duty, they provide assistance in (relatively uncommon) gynecological emergencies, occasionally monitor mothers' health after birth, and are present in the labor room when decisions regarding a delivery are made, or if complications arise. Midwives, on the other hand, work 12-hour shifts (from 8 am to 8 pm).

For all of the hospitals in our sample, there are at least two obstetricians and two midwives on duty during the night, and each doctor assists on average between 1 and 2 deliveries per night. During these times, each delivery thus accounts for a major part of a doctor's duties. Although in our setting doctors cannot leave the hospital while they are on duty, beds are available to rest when there is no emergency or complication that requires their presence (Ministerio de Sanidad y Política Social, 2009).

1.3 DATA AND METHODS

1.3.1 *Description of the data*

Our data consists of all 6,163 birth records from four public hospitals in different Autonomous Regions in Spain during the years 2014-2016[4]. The characteristics of the hospitals in our sample are comparable to that of the majority of public hospitals in Spain, in particular with regard to the volume of births attended per year (between 300 and 1500). In terms of c-section rates, three of the four hospitals are in the left tail of the distribution, while one is just at the mode, with a c-section rate around 21%. This comparison can be found in figure 1.A.1 in the appendix.

Each birth registry contains information on the mother's characteristics (age, nationality, education, marital status, etc.), on the pregnancy, on the type of birth (planned cesarean, unscheduled cesarean, eutocic delivery, etc.), on medical interventions during labor, on a series of medical indicators collected before, during, and after the delivery, on the newborn (birth weight, Apgar scores, etc.), and on the date and time of birth. Table 1.A.1 shows some summary statistics of the variables of interest[5]. In our data, 5% of women delivered via a planned c-section, more than 11% via an unplanned c-section, and 68% had an eutocic delivery, that is, a vaginal delivery without other interventions

[4] Data collection was approved and financed by the Spanish Ministry of Health under the Strategy for Assistance at Normal Childbirth in the National Health System (PI/01445).

[5] For comparison, in table 1.A.2 we show descriptive statistics of the coincident variables reported in the Spanish National Statistics Institute birth registries for all births that took place in hospitals in Spain in the years 2014-2015. We see a slightly higher proportion of non-Spanish women in our data and also less multiple pregnancies, but similar characteristics in terms of age, gestational length or birth weight.

(i.e. spatula, forceps, or vacuum). Vaginal deliveries with such interventions represent around 15% of the sample. We eliminate non-single births, planned c-sections and breech vaginal babies[6]: our final sample consists of 5,783 observations.

Our main outcome variables are Apgar scores at minutes 1 and 5 after birth. These result from the examination of the health status of the newborn performed by the midwife or the pediatrician one and five minutes after birth, respectively (AEPED, 2014)[7]. In particular, they assess and grade between 0 and 2 points each of the following aspects: appearance (skin color), pulse (heart rate), grimace (reflex irritability), activity (muscle tone), and respiration. These variables thus take values between 0 and 10. We study both the levels of these scores and also the probability of the scores being below different thresholds. We also look at whether the newborn needed reanimation (assisted ventilation), whether they were admitted to the intensive care unit, and at the event of neonatal death.

Some other medical variables included in our analysis need further clarification. Besides the outcome variables presented above, another one of interest is the umbilical cord pH, which is an indicator of fetal distress. A sample of blood from the umbilical cord artery is collected after cord clamping, and the levels of pH are measured. There is some variation in the literature in what is considered the range of normal values for this outcome, with thresholds for acidemia (low pH) spanning from 7 to 7.20 (Malin et al., 2010). In our analysis we consider thresholds of 7.20, 7.15, and 7.10. A related variable is the fetal scalp pH

[6]Breech vaginal babies – that is, babies that were in breech position and were born vaginally – are a rare case: we only have 8 of those in our sample. This is because attending such type of birth requires special caution and expertise (American College of Obstetricians and Gynecologists, 2006) – most fetus in breech position are delivered by planned c-section. Therefore, these kind of births are not a plausible counterfactual for unplanned cesareans.

[7] In general, Apgar scores can be determined by a pediatrician, a midwife or a nurse present in the labor room – this depends mainly on the routines of each hospital. In the hospitals in our sample, this task is normally assigned to midwives.

or intrapartum pH, which is a measure of fetal distress during labor, before birth. In this case, the pH is measured from a sample collected from the baby's head when it becomes visible. Too low values of this variable – in particular, pH lower than 7.20 – suggest that the baby is not getting enough oxygen, and thus a cesarean section might be necessary (SEGO, 2005). Finally, one relevant control we include in our preferred specifications is obstetric risk. This is recorded by the medical professionals who prepared our data, and defined as a dummy variable that takes value one if, during pregnancy, some risk factors were detected that could lead to an adverse pregnancy outcome[8].

1.3.2 *Variation in the c-section rate by time of day*

Figure 1.1 shows the c-section rate at different times of day for our sample of public hospitals in Spain. We can observe that the distribution of unscheduled c-sections by time of birth is not uniform. The proportion of women that deliver via an unplanned c-section is higher in the early hours of the night (from 11 pm to 4 am), and much lower during the remaining hours of the night and the rest of the day. This pattern is not matched by either the total number of births or the number of vaginal births (see figure 1.A.2 in the appendix). More importantly, this variation is not driven by differences in maternal or pregnancy characteristics of the deliveries that take place at different times of day. In the next section, Table 1.1 confirms the balance of a very large set of mother and pregnancy characteristics between women delivering in the early hours of the night and during the rest of the day. As we will

[8] More specifically, obstetric risk was defined as the presence during pregnancy of one or more of the following factors that increase the chance of an adverse pregnancy outcome: cholestasis, chorioamnionitis, 486 diabetes insulin and non-insulin dependent, chronologically prolonged pregnancy, multiple pregnancy, hellp syndrome, hypertension, isoimmunization in pregnancy, stained amniotic fluid, fetal malformation, uterine malformation, fetal malposition, myomectomy, oligoamnios, previous preterm labor, placenta praevia, plyhydramnios, preeclampsia, premature rupture of membranes, siphylis, toxoplasmosis, previous c-section, repeated abortions, previous miscarriages, anteparturm alteration of fetal wellbeing.

discuss in further detail, this allows us to use this exogenous variation as an instrument for mode of delivery.

We are not the first to document this early night spike in unscheduled c-section deliveries. For example, Fraser et al. (1987), Brown (1996), and Spetz et al. (2001) show an increase in the probability of a c-section at the end of the day up until midnight, and Hueston et al. (1996) documents a peak in the unplanned c-section rate between 9 pm and 3 am. These authors have interpreted these evening or night peaks as evidence that convenience and doctors' demand for leisure influence the timing and mode of delivery. Similarly, several studies find that the probability of a c-section increases when doctors can go to sleep or return home after the birth, likely linked to the fact that cesarean sections require on average less total time devoted to the patient (Klasko et al., 1995; Spong et al., 2012).

This explanation is consistent with the time pattern that we observe in our data. Given the medical shift structure and the larger time-cost of surveillance implied by vaginal deliveries, doctors' incentives to perform c-sections in ambiguous cases may vary by time of day. In particular, we expect doctors to have a larger incentive to perform c-sections in the early hours of the night. By this time, on-duty doctors have already been working for more than 12 straight hours (see Figure 1.A.3 in the appendix[9]). If they perform a c-section and do not have other mothers to care for, they can expect to rest for the remainder of their shift. Alternatively, if they do not perform a c-section, they will need to occasionally monitor the vaginal delivery throughout the night. Moreover, ongoing deliveries in the early hours of the night have a high probability of falling under the responsibility of the doctor

[9] Figure 1.A.3 shows the proportion of unplanned c-sections as a function of the number of hours worked by physicians: 0 hours corresponds to 8 am. As can be seen, the proportion of c-sections starts to increase when doctors have been working for already 12 hours, and reaches its maximum when hours worked are between 15 and 20. The proportion of unplanned c-section decreases in the last hours of their shift.

on duty[10], as opposed to deliveries which begin later and are more likely to finish past the doctor's shift. These conditions would suggest that a higher share of deliveries with ambiguous indications end up as cesarean sections during the early hours of the night, as compared to the rest of the day. Consistent with this interpretation, we find that the probability of doctors performing a c-section at these times increases when there is only one ongoing delivery at the beginning of the night, that is, when the expected marginal gain of a c-section is larger[11].

Other alternative explanations are not compatible with this variation. For example, if either patient's or physician's fatigue increased the probability of c-sections, we would expect to see a higher unplanned c-section rate during the late hours rather than the early hours of the night. We can also rule out that this is driven by an accumulation of births during these hours, as we do not observe the same time pattern for the number of births (see figure 1.A.2 in the appendix). Finally, the early night spike in c-sections cannot be explained by selection of highly interventionist doctors at different times of day, as deliveries are not pre-assigned to a given obstetrician. We also provide evidence that this is not the case in Figure 1.A.4 in the appendix[12], where we show that there are no systematic differences among doctors in the probability of attending births during the early hours of the night.

1.3.3 *Identification strategy*

Our objective is to identify the causal impact of non-medically indicated c-sections on infants' health at birth. The simple comparison

[10] Average duration for the first stage of labor in vaginal deliveries among first-time mothers is around 8 hours (NICE, 2014), and for the second stage around 3 hours. See section 1.2.1 for more detail.

[11] Table 1.A.3 in the appendix shows that the increase in the probability of cesarean birth at the early hours of the night (from 11 pm to 4 am) is larger in days when there is only one birth at night compared to days with more than one birth.

[12] Figure 1.A.4 plots, for a small sample of births for which we know the doctor who attended the delivery, the probability of attending births during the early hours of the night across different doctors.

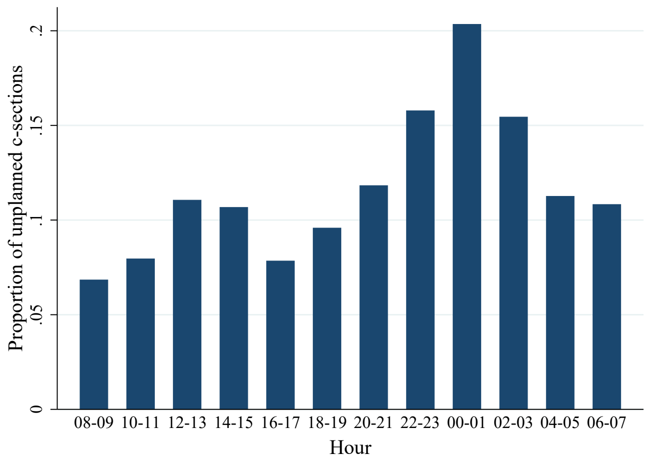


FIGURE 1.1: Proportion of Unplanned C-sections by Time of Day

Notes: The figure represents the proportion of unplanned c-sections by time of day over the sample of unplanned c-sections and vaginal births. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

of women who had a c-section and those who delivered vaginally is likely to suffer from omitted variable bias, as these groups likely differ in characteristics that influence the outcome variables. Table 1.A.4 in the appendix compares observable characteristics of these two types of mothers. We observe, in fact, that these mothers are significantly different in terms of several relevant aspects such as age, gestational length, obstetric risk, or educational achievement, all potentially related to the health of the newborn. There are thus reasons to be concerned that they might also differ in other characteristics we cannot observe. Moreover, a comparison of vaginal deliveries and births by c-section does not allow to identify which kind of c-section is causing whatever health effects are found, since we observe the outcomes of both medically and non-medically indicated interventions. In order to overcome these issues, we use variation in the probability of having a c-section by time of day. The purpose of the instrument is thus twofold: to compare similar women, and to precisely identify the impact of non-medically indicated cesareans.

We define a binary variable CS_i equal to one if the mode of delivery is an unplanned c-section and zero if it is a vaginal delivery (eutocic or operative). Infant health H_i refers to either Apgar scores or other measures of neonatal health. We would thus like to estimate the following equation:

$$H_i = \beta_0 + \beta_1 CS_i + \beta_2 X_i + \varepsilon_i \tag{1.1}$$

where X_i is a set of covariates that include information on mothers' personal and pregnancy characteristics. As discussed earlier, the estimation of equation (1.1) is, however, likely to provide biased estimates of β_1. To overcome this potential endogeneity, we use an IV approach, instrumenting the type of birth with an indicator for the time of day the infant is born. Therefore, our first stage is as follows:

$$CS_i = \gamma_0 + \gamma_1 earlynight_i + \gamma_2 X_i + \upsilon_i \tag{1.2}$$

where $earlynight_i$ is an indicator variable equal to 1 if woman i gives birth during the beginning of the night (from 11 pm to 4 am). We expect a positive $\widehat{\gamma}_1$ since obstetricians are more likely to initiate a c-section during these hours of the night in order to gain time for rest or leisure.

The identifying assumption is that $earlynight_i$ is not correlated with ε_i, but this assumption entails two conditions. The first is that the instrument is as good as randomly assigned. We provide suggestive evidence that this is the case by comparing personal and pregnancy characteristics of mothers who give birth between 11 pm and 4 am and those during the rest of the day in Table 1.1. Mothers are similar with respect to their age, educational level, weight and height, alcohol and tobacco consumption habits during pregnancy, gestational length, obstetric risk, weight of the newborn, or previous c-sections. The level of intrapartum pH, a measure of fetal distress during labor – a major cause of emergency c-sections – is also equivalent. Mothers are also comparable in terms of the average time that they have been in the hospital, that is, time between admission and time of birth. We find some slight differences between mothers across time of day with respect to nationality (there are slightly more non-Spanish women during the day shift) and marital status (more unmarried women during the day). However, these differences are very small in magnitude. We also find that the proportion of women whose labor was induced is higher during the early hours of the night (28.5%) compared to the rest of the day (22.6%). This is something one might expect from our institutional setting, since in the hospitals in our sample most inductions are performed in the morning and, given the average duration of labor, these women are more likely to give birth during the early hours of the night. We control in our main specification for all of these differences and perform a robustness check excluding inductions in Section 1.5.2, where we find that our conclusions still hold. Overall, we thus feel confident with the assumption that there is no selection of women into the different times that could threaten our identification.

Additionally, identification requires the exclusion restriction to hold; that is, the instrument should affect infant health only through the increased probability of having a c-section. One potential concern is that the quality of medical care could change depending on the time/shift. Although we do not have a direct measure of hospital service quality, we have some information about the doctors attending the birth for a subsample of births. In table 1.A.5 we show that the number of doctors and the proportion of male doctors is balanced across different times of day. Additionally, we provide more systematic evidence in favor of our exclusion restriction by performing the analysis using variation in the probability of having a c-section only during the night, thus holding the quality of medical care constant (see section 1.5.1).

Table 1.1: Maternal Characteristics by Time at Delivery

	Means		p-value
	Rest of the day	Early night	for difference
A. Personal characteristics			
Mother's age	31.729	31.888	0.349
Level of education			
No school	0.033	0.025	0.146
Primary school	0.254	0.262	0.563
Secondary school	0.525	0.523	0.906
University education	0.187	0.189	0.876
Non-Spanish	0.256	0.223	0.015
Single	0.019	0.009	0.017
Mother's weight	65.561	65.779	0.630
Mother's height	1.650	1.607	0.534
B. Pregnancy characteristics			
Tobacco during pregnancy	0.120	0.126	0.606
Alcohol during pregnancy	0.004	0.004	0.891
Gestation weeks	39.263	39.274	0.853
Previous c-section	0.090	0.103	0.173
Obstetric Risk	0.388	0.409	0.161
Intrapartum pH*	7.271	7.278	0.402
Birth weight	3277.356	3270.303	0.662
Induction	0.226	0.285	0.000
Time in hospital (in hours)*	9.891	10.156	0.450
Observations	4478	1305	5783

Notes: The table shows means for a set of maternal and pregnancy characteristics by time of day and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies). Variables marked with an asterisk (*) are not available for the whole sample. Intrapartum pH is only available for a sample of births (425 observations), and time in hospital is only available for one hospital (2289 observations).

1.4 RESULTS

Tables 1.2 and 1.3 present the results for the OLS estimation of equation (1.1) for the different measures of neonatal health. In table 1.2, the first column for each outcome presents the results without controls, the second column incorporates controls for maternal characteristics, and finally the third column adds information about the pregnancy. All specifications include hospital and weekday fixed effects, the sample is restricted to single births, unplanned c-sections and vaginal deliveries, and we cluster standard errors at the hospital-shift level[13]. The results show that delivering via a c-section is associated with a significant decline of Apgar scores 1 and 5. Table 1.3 presents the results for other outcomes of neonatal health. As it can be seen, babies born by cesarean section are more likely to need reanimation and to go to the intensive care unit, but they are no more likely to die.

As explained above, these estimates are likely to be biased because mothers giving birth by c-section and vaginally are not comparable, and because we cannot identify which kind of c-section is driving the results. The results for the IV estimation of the effects of non-medically indicated c-sections on Apgar scores 1 and 5 are shown in Table 1.4[14]. The first stage F-statistics are larger than 34 for the different specifications, so following Stock and Yogo (2005) critical values with one endogenous variable and one IV (16.38), we can reject the null hypothesis that our instrument is weak. In line with our descriptive analysis, Panel B shows that births that take place between 11 pm and 4 am are around 6 percentage points more likely to be by cesarean[15].

[13] All estimations hereafter use clustered standard errors at the hospital-shift level. We show in Table 1.A.6 in the appendix that our IV results are robust to alternative standard error estimations.

[14] The full regression output for both the first and second stage can be found in tables 1.A.7 and 1.A.8 in the appendix.

[15] We have also considered alternative specifications of the IV, using dummies for single hours in the window from 11 pm to 4 am. Our second stage results are similar but the first stage is weaker, thus harming precision and raising concerns about bias of

Table 1.2: OLS Results – Apgar Scores

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Unplanned CS	-0.528***	-0.524***	-0.419***	-0.219***	-0.219***	-0.142***
	(0.057)	(0.057)	(0.061)	(0.038)	(0.037)	(0.043)
Mean of Y		8.895			9.798	
Observations		5783			5781	
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the results of OLS regressions of Apgar scores 1 and 5, respectively, on an indicator for an unplanned cesarean birth. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3: OLS Results – Other Outcomes

	Intensive Care Unit		Reanimation		Neonatal death	
	(1)	(2)	(3)	(4)	(5)	(6)
Unplanned CS	0.137***	0.102***	0.081***	0.062***	-0.001	-0.005
	(0.016)	(0.014)	(0.014)	(0.014)	(0.002)	(0.003)
Mean of Y	0.060		0.082		0.004	
Observations	5783		5782		5783	
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls		✓		✓		✓

Notes: The table shows the results of OLS regressions of different indicators of neonatal health on an indicator for an unplanned cesarean birth. The outcome variable in columns (1)-(2) is a dummy variable equal to one if the newborn was admitted to the intensive care unit; in columns (3)-(4), an indicator for whether the newborn needed reanimation (assisted ventilation), and in columns (5)-(6) an indicator of neonatal death. The first column for each outcome shows the results of this regression controlling for maternal characteristics, weekday and hospital fixed effects; in the second column pregnancy controls are also added. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In the first row of the table below (Panel A), we observe that a c-section has a negative impact on both Apgar score 1 and Apgar score 5. The estimated effects are large and significant. In the specification with the full set of controls (column 3), an unscheduled c-section reduces Apgar score 1 by 0.992 points. This effect is around 0.9 standard deviations (1.117) and is significant at the 10% significance level. A c-section also has a negative impact on Apgar score 5. In this case the coefficient is -0.936, larger than one standard deviation (0.818) and significant at the 5% significance level.

Most of the newborns in our sample have an Apgar score 1 equal to 9 and an Apgar score 5 equal to 10 (see figure 1.A.5). We thus perform a similar analysis but using as dependent variable an indicator for having Apgar scores 1 and 5, respectively, lower than 10 (table 1.A.9), and both scores lower than 9 (table 1.A.10). Our qualitative conclusions hold, as we find that a non-medically justified c-section, as compared to a vaginal delivery, increases the probability of having Apgar scores 1 and 5, respectively, below 10 by around 25 and 40 percentage points, and the probability of having Apgar scores 1 and 5 below 9 by 36 and 19 percentage points. Finally, Figure 1.A.6 in the appendix provides an overview of the size of the coefficients for different thresholds of Apgar 1 and 5, respectively, as dependent variables. This is relevant, since decreases in Apgar scores are non-linearly related to the health of the newborn. We see a clearer pattern for Apgar scores 5: there seems to be an effect of these non-medically justified interventions on the probability of having Apgar scores lower than 10, 9 and 8, but not lower than 7 or inferior levels. Therefore, these marginal c-sections increase the probability of deviating from the perfect scores, which are the mode in our sample, but we do not see significant effects in the left tail of the distribution.

We also perform the same analysis for other infant health outcomes. Results can be found in Table 1.5. Although we might expect an effect

the 2SLS. Results are available upon request.

Table 1.4: IV Estimation – Apgar Scores

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.122**	-1.147**	-0.992*	-0.956**	-0.987**	-0.936**
	(0.497)	(0.501)	(0.572)	(0.404)	(0.408)	(0.464)
Mean of Y		8.895			9.798	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned c-section on Apgar scores 1 and 5, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

on needing intensive care, reanimation, or neonatal mortality, we do not observe any significant impact.

Our IV identifies the local average treatment effect for the "marginal" women, that is, for the deliveries that are sensitive to the subjective assessment of the doctor. More specifically, we capture cases in which the time of birth affects the decision of the doctor to perform a cesarean section. We therefore focus on c-sections that are not strictly necessary in the medical sense and that are potentially avoidable surgeries. These are, in fact, arguably the most relevant from a policy point of view.

Table 1.5: IV Estimation – Other Outcomes

	Intensive Care Unit		Reanimation		Neonatal death	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. 2SLS						
Unplanned CS	0.154	0.092	0.101	0.057	0.030	0.026
	(0.103)	(0.114)	(0.114)	(0.133)	(0.031)	(0.035)
Mean of Y	0.060		0.082		0.004	
Panel B. First stage						
Early night	0.073***	0.063***	0.073***	0.063***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5782	5782	5783	5783
First-stage F	41.591	34.234	41.576	34.149	41.591	34.234
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls		✓		✓		✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on different indicators of neonatal health. The outcome variable in columns (1)-(2) is a dummy variable equal to one if the newborn was admitted to the intensive care unit; in columns (3)-(4), an indicator for whether the newborn needed reanimation (assisted ventilation), and in columns (5)-(6) an indicator of neonatal death. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling for maternal characteristics, weekday and hospital fixed effects; in the second column pregnancy controls are also added. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We are not able to estimate the effect for women who have a clear indication for a vaginal delivery or for women who receive c-sections that are medically indicated.

If we compare the results from the IV and OLS estimations, the IV coefficients are larger in absolute terms for Apgar scores. This can be explained by the fact that with the OLS estimation we include medically indicated c-sections, which reduce fetal distress and this partially offsets the negative effects of the non-medically indicated c-sections that we find when using our instrument.

However, if we compare the results for the other outcomes (see tables 1.3 and 1.5), we observe that in this case OLS coefficients are larger and significant: c-sections are associated with an increased probability of needing intensive care and reanimation. This suggests that these medically-indicated c-sections are performed in order to assist infants in distress who need immediate support. On the other hand, the IV estimates are not significant, arguably because the effects of non-medically indicated c-sections are short-lived: in spite of the worsening in Apgar scores, we do not find substantial evidence that these negative effects translate into needing intensive care, reanimation, or increased mortality risk.

To support the interpretation that our IV identifies the effect of non-medically indicated c-sections, we provide evidence that the c-sections captured by our instrument are not correlated with indications that should predict a medically necessary cesarean. In particular, we show that, while unplanned c-sections are in general strongly correlated with fetal distress, as measured by the level of intrapartum pH, we do not see any relationship when we focus on the predicted c-sections from our first stage. This comparison can be found in table 1.A.11 in the appendix.

So far, our analysis has compared c-sections with all vaginal births. The latter comprise two main categories: eutocic births – without any instrumentation – and operative (or instrumented) vaginal deliveries, which involve the use of forceps, vacuum or spatula. Medical studies have documented a negative association between operative vaginal deliveries and infant health (American College of Obstetricians and Gynecologists, 2015). Moreover, the decision to perform these procedures is also subject to variation at the provider level (Webb, 2002). For a cleaner comparison without the potential manipulation of the control group, we perform the same analysis comparing c-sections with eutocic deliveries. We would expect the effects of non-medically indicated c-sections to be stronger if compared with this group. The results in table 1.A.12 seem to confirm this hypothesis, and we also observe a slightly stronger first stage, suggesting that physician impatience might also lead to an increased use of instrumentation in the early hours of the night.

1.5 Robustness Checks and Extensions

1.5.1 *Exclusion restriction: variation within the night*

One potential concern of our identification strategy is that the quality of medical care could differ during the day compared to the night. Hence, it may be that the negative effects that we find on infant health are not due to the increased probability of having a c-section, but rather to a reduction in the quality of care during this time.

To further investigate this issue, we perform the same IV estimation but restricting the sample to mothers who gave birth during the night. We thus use variation in the probability of having a c-section during the night, holding the quality of care constant. As before, our instrument is an indicator variable equal to 1 if the woman gives birth during the early hours of the night (from 11 pm to 4 am). The sample is restricted to deliveries taking place from 8 pm to 8 am; i.e., during the last half of physicians' shifts, when healthcare professionals in the labor room – both obstetricians and midwives – do not change.

Results for the IV estimation using variation during the night can be found in Table 1.6. Despite the smaller sample size, we again find that a c-section reduces both Apgar scores 1 and 5. The coefficients remain large and significant, in particular so for Apgar 5. We interpret these results as evidence in favor of our exclusion restriction.

Table 1.6: IV Estimation – Apgar Scores during the Night

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.530*	-1.524*	-1.413	-1.511**	-1.512**	-1.535**
	(0.814)	(0.830)	(0.964)	(0.653)	(0.663)	(0.766)
Mean of Y		8.879			9.790	
Panel B. First stage						
Early Night	0.054***	0.053***	0.044***	0.053***	0.053***	0.044***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	3023	3023	3023	3022	3022	3022
First-stage F	17.217	16.619	12.812	17.144	16.537	12.760
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, for births that took place between 8 pm and 8 am. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) that took place during the night. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1.5.2 *Excluding inductions*

The comparison of maternal characteristics in Table 1.1 showed that mothers giving birth in the early hours of the night are more likely to have had their labor induced. Inductions can be scheduled, typically because the pregnancy has gone beyond full term and labor has not spontaneously started, or can be unscheduled if the mother's waters break but labor does not begin (NICE, 2008). If an induction is to be scheduled, the hospitals in our sample usually plan the latter for the morning, such that after progression of labor at average pace these women are expected to give birth in the evening or during the early hours of the night.

The relation between inductions and c-sections is a question where the medical literature and medical practice seem to differ. We observe in our sample that mothers with induced labor are more likely to have a c-section (see table 1.A.4). However, the recent medical literature finds that, while c-sections are conventionally regarded as the main potential complication of inductions, inductions at full term do not increase the risk of cesarean delivery (Saccone and Berghella, 2015) or even lower it (Mishanina et al., 2014), with no increased risks for the mother and some benefits for the fetus. All in all, it seems that whether or not a c-section is needed in cases of induced labor is likely to be dependent on the assessment of the obstetrician, such that mothers having had inductions probably fall into a "gray area" where we expect doctors' decisions to be more sensitive to external factors and incentives.

In any case, even if the decision to perform a c-section on mothers with induced labor was more dependent on doctors' routines or incentives than on the health conditions of the mother and the baby, if our analysis was driven by this type of mother alone, we would not be able to disentangle the effect of c-sections from the effect of medical inductions. In our main specifications we directly control for whether labor was induced, but in Table 1.7 we also repeat our analysis excluding inductions from our sample[16]. Here we see that, despite the reduction in the number of observations, our qualitative conclusions hold: births in the early night are still more likely to end up as cesarean sections, and these have a negative and significant impact on Apgar scores. We thus conclude that, although inductions seem to make our first stage stronger as they might offer room for discretionary behavior, our findings do not depend on including them.

[16]The results for both the specification without inductions and the specification with only births during the night for reanimation, ICU admission, and neonatal death are consistent with those of table 1.5. Results are available upon request.

Table 1.7: Robustness Check – Excluding Inductions

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.747	-1.769	-1.804	-1.804*	-1.847*	-1.921*
	(1.086)	(1.104)	(1.171)	(0.931)	(0.952)	(1.011)
Mean of Y		8.952			9.828	
Panel B. First stage						
Early Night	0.037***	0.037***	0.035***	0.037***	0.037***	0.035***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	4369	4369	4369	4367	4367	4367
First-stage F	10.720	10.663	10.179	10.677	10.614	10.319
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, for non-induced births. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, and an indicator for preterm birth. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) that were not induced. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1.5.3 *Falsification test*

In order to lend support to the credibility of our identification strategy, we run additional "placebo" regressions using an outcome variable that is predetermined when the mother goes into labor, and thus should not be affected by a c-section. In particular, we analyze birth weight and weeks of gestation. The results of this analysis are reported in Table 1.8. As in previous tables, the first column for each outcome presents the results without controls, the second column incorporates controls for maternal characteristics, and finally the third column adds information about the pregnancy. The results of this exercise suggest that there is no effect of c-sections on birth weight or gestational weeks. This provides further evidence in favor of our specification.

Table 1.8: Placebo Regressions: Birth Weight and Gestational Weeks

	Birth Weight (in logs)			Gestational weeks		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-0.023	-0.027	0.042	0.250	0.203	0.081
	(0.077)	(0.076)	(0.077)	(0.774)	(0.772)	(0.866)
Mean of Y		8.080			39.266	
Observations	5782	5782	5782	5783	5783	5783
First-stage F	41.627	41.559	34.222	41.661	41.591	35.154
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on birth weight (in natural logs) and gestational weeks, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth (except in the regression of gestational weeks), and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1.5.4 *Time of admission and time of birth*

One potential concern with using time of birth as an instrument for the mode of delivery is that, given that cesarean sections by definition shorten labor, the exact time of birth will be influenced by the type of birth itself. In other words, one might be worried about reverse causality in the first stage. We argue that any potential bias should be alleviated by the specification of the instrument not as the time of birth itself, but as a relatively wide time interval (in particular, as a dummy equal to one for births between 11 pm and 4 am). Because the instrument is defined in this way, we do not need to assume that the exact time of birth is not influenced by the mode of delivery; it suffices that any impact of the decision about the type of birth on the time interval in which the delivery takes place is negligible.

In our context, if doctors' incentive is to perform a cesarean section to ongoing deliveries early at night that they expect to end up during their shift, it will likely be to mothers that are advanced in labor. Therefore, the counterfactual to the cesarean is expected to be a vaginal birth two or three hours later[17]; that is, for most c-sections in the early night, the counterfactual vaginal birth would have probably taken place in the early hours of the night as well. As a result, the change in the probability of giving birth between 11 pm and 4 am caused by having a c-section is expected to be small.

In order to assess empirically the magnitude of the potential bias, we use information about the time of admission of mothers to the hospital, which is only available for one of the hospitals in our sample. In particular, we want to see if our results are robust to substituting our instrument with one based on the time of admission. This alternative instrument should remove concerns about reverse causality since, for unscheduled deliveries, time of admission should not be affected by mode of delivery.

First, we explore the distribution of the c-section rate as a function of time of admission (see figure 1.A.7) and find that there is a similar peak to that in figure 1.1, in this case for mothers admitted between 2 pm and 8 pm. Therefore, we define our new instrument to be equal to one for mothers admitted during this time interval[18]. Results using this

[17]See an explanation of the average time of each stage of labor in section 1.2.1.

[18]Following the same logic as in our main analysis, we select the interval in which

new instrument can be found in table 1.9, which follows the usual table structure. Panel B displays the coefficients of the first-stage regressions: in the third column for each outcome, which shows the results of the specification with the full set of controls, we can see that mothers that arrived at the hospital between 2 pm and 8 pm were around 6.3 percentage points more likely to have a c-section. This is the same result we found for mothers giving birth between 11 pm and 4 am: they are also 6.3 percentage points more likely to have a cesarean birth. Panel A shows the 2SLS coefficients: despite the reduced sample size, we find very similar point estimates to those in table 1.4. The resemblance of these results to those in our main analysis suggests that reverse causality, in practice, does not have a large influence in our setting, and supports the validity of our instrument.

1.5.5 *Another measure of neonatal health: umbilical cord pH*

In addition to Apgar scores, reanimation, ICU admission and neonatal death, we also study the impact of cesarean sections on the pH of the umbilical cord. Although it has not been used in the economics literature, this measure of neonatal health has been widely analyzed in medical studies, and it is considered to add objective information to the Apgar score regarding the status of the newborn. Due to its objective nature, it is used to support medico-legal claims (Skiold et al., 2017). As explained in Section 1.3.1, the examination of the umbilical artery provides a measure of fetal distress. Although the relationship between pH levels and Apgar scores is not one-to-one, they are positively correlated[19]. The medical literature recommendation is to consider pH

the c-section rate is above 15%.

[19]Figure 1.A.8 in the appendix shows the distributions of umbilical cord pH for infants with Apgar scores 1 above and below 9 (first panel), and for infants with Apgar scores 5 above and below 9 (second panel). We observe that the distribution of pH levels for infants with Apgar scores below 9 is shifted to the left compared to that for babies with higher scores, with this being more salient for Apgar score 5.

Table 1.9: Robustness check – IV Estimation with Admission Time Instrument

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.554**	-1.568*	-1.601*	-0.802	-0.791	-0.793
	(0.787)	(0.815)	(0.960)	(0.578)	(0.601)	(0.712)
Mean of Y		8.861			9.869	
Panel B. First stage						
Admission time 2pm-8pm	0.077***	0.074***	0.064***	0.077***	0.074***	0.063***
	(0.022)	(0.022)	(0.021)	(0.022)	(0.022)	(0.021)
Observations	2289	2289	2289	2287	2287	2287
First-stage F	12.079	11.601	9.465	12.029	11.550	9.423
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned c-section on Apgar scores 1 and 5, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for mothers admitted to the hospital between 2 pm and 8 pm. Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

levels together with Apgar scores in order to assess the well-being of the newborn (Hannah, 1989; Malin et al., 2010).

Table 1.10 shows the results from the estimation of the impact of a c-section on the probability of the pH level being below different thresholds (7.20, 7.15 and 7.10) for the different samples: the full specification (columns 1–3), during the night (columns 4–6) and excluding inductions (7–9). This outcome was only recorded in 3 out of the 4 hospitals in our sample, and thus the number of observations is lower. All our estimates go in the same direction: c-sections increase the probability of pH levels being below the different thresholds, suggesting the presence of a negative health effect as measured by this outcome. The most consistent results are found for the pH threshold of 7.15. Our first stage F-statistic is strong for the full specification (25.58) but becomes weaker as the sample drops. Overall, these findings go in line with the previous results of a negative effect of c-sections on neonatal health.

Table 1.10: IV estimation — Umbilical cord pH level

	Full Specification			During the Night			Excluding Inductions		
pH threshold	7.20	7.15	7.10	7.20	7.15	7.10	7.20	7.15	7.10
Panel A. 2SLS									
Unplanned CS	0.303	0.341*	0.184	1.074*	0.857**	0.307	1.004	0.947*	0.573*
	(0.250)	(0.192)	(0.122)	(0.562)	(0.415)	(0.220)	(0.671)	(0.538)	(0.333)
Mean of Y	0.221	0.102	0.042	0.212	0.100	0.044	0.216	0.096	0.039
Panel B. First stage									
Early Night		0.063***			0.042***			0.033***	
		(0.012)			(0.014)			(0.012)	
Observations		4444			2316			3403	
First-stage F		25.589			8.567			6.992	

Notes: The table shows the instrumental variable estimates of the effect of an unplanned cesarean birth on the probability of the umbilical cord pH being below different thresholds (7.20, 7.15, and 7.10), for different samples. Columns (1)-(3) use the usual full sample, columns (4)-(6) use only births during the night, and columns (7)-(9) include only non-induced births. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor (except in the last three columns). Mean of Y refers to the average of the outcome variable in the sample. The sample is in all cases restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1.6 CONCLUSIONS

This paper provides new credible evidence of the adverse effects of avoidable cesarean sections on newborn health. In order to overcome potential omitted variable bias and abstract from those cases in which c-sections respond to a clear clinical indication, we make use of a novel instrument that exploits variation in the probability of receiving a c-section that is unrelated to maternal and fetal health: variation in time of birth. Specifically, we document an increase in unplanned c-sections during the early hours of the night (from 11 pm to 4 am) that is not driven by different characteristics of mothers who give birth during this time, providing us with exogenous variation in the probability of the delivery ending up in a cesarean section.

Our findings suggest that these non-medically indicated c-sections lead to a significant worsening of newborn health, as measured by Apgar scores. According to the medical literature, deterioration in these outcomes might be capturing increased respiratory problems and reduced excitability and muscle tone (Hyde et al., 2012). However, the magnitude of our estimates suggests that these c-sections lead to a decrease of just around one point in Apgar scores 1 and 5 in otherwise healthy babies – the mean Apgar scores 1 and 5 are 8.9 and 9.8, respectively. Our analysis by thresholds of Apgar scores confirms that the effects of these c-sections are limited to the higher levels of these scales; in particular, we see an increased probability of having Apgar score 5 below 10, 9 and 8. It is worth noting that previous studies find worse long-run outcomes for newborns with these levels of Apgar, compared to their siblings with perfect scores, even if these levels are not generally considered to be concerning: Oreopoulos et al. (2008) find that individuals with Apgar scores of 7 or 8 are more likely to drop out or repeat a grade, and that those with Apgar scores between 7 and 9 are also more likely to receive social assistance after age 18.

In any case, we do not find evidence that these effects translate into a significant increase in the need for reanimation or intensive care, or into increased risk of neonatal death, which is consistent with the absence of significant impacts on lower levels of Apgar scores and on low thresholds of the pH of the umbilical cord. We can thus rule out very severe impacts at birth, as well as any short-run health benefit of these avoidable interventions. This is an important contribution, given that previous studies in the medical literature documented an association between c-sections and an increased risk of serious respiratory morbidity and subsequent admission to neonatal ICU (Grivell and Dodd, 2011). Their findings are consistent with the results of our OLS estimation, suggesting that former analysis might have been capturing the underlying health status of newborns who need a medically necessary cesarean.

However, it should also be pointed out that some effects of c-sections may not be visible at birth. In particular, medical studies suggest that the exposure of newborns to the maternal vaginal microbiota is interrupted with cesarean birthing, and that this could translate into increased risk for immune and metabolic disorders in the long run (Hyde et al., 2012; Dominguez-Bello et al., 2016). Any such effect need not be reflected in any of the short-run outcomes we are able to explore in this study, which limits the conclusions we can derive from our analysis. In this paper, however, we propose a new instrument that will make possible to examine this and other channels and gather evidence to obtain a more complete understanding of the causal effect of non-medically indicated c-sections on the health of the infant and the mother in the longer run.

Our results also highlight non-financial incentives as an important factor influencing the decision-making of health care providers. Although more work is needed to clearly understand the decisions of doctors driving the observed time variation in c-section rates, we have provided some suggestive evidence that stresses the potential role of leisure incentives in the context of public hospitals, and which is consistent with the findings of previous studies. In particular, our findings suggest that doctors may be less tolerant to the time-consuming natural progression of labor during times of day when leisure incentives are more salient, and thus are more willing to perform procedures that accelerate the delivery. Along this line, our results point to the need to revise the incentives created by the shift structure and long working hours of physicians, so as to reduce avoidable interventions.

A simple back-of-the-envolope calculation can shed some light on the potential gains that could result from such reduction. The first-stage coefficient from our main specification with all controls (column 3 in table 1.4) implies that, holding all other characteristics constant, during the early hours of the night the c-section rate increases by 6.3 percentage points compared to the rest of the day. Given that the c-section rate in our sample of hospitals is 16.5%, removing these excess c-sections would lower the c-section rate by 38.1% – or equivalently, a decrease of 245 c-sections per year[20]. Taking into account that the average cost of a c-section for the Spanish public health system is 1692.97 Euros higher than that of a vaginal delivery[21], by cutting these excessive c-sections, hospitals in our sample could achieve a cost reduction of around 675,500 Euros. Applying the same logic for all births that took place in Spanish public hospitals in 2014, this would result in savings of more than 47 million Euros for the Spanish health system[22]. To give some meaning to these numbers, given that the average annual salary for a speciality doctor is 45,970 Euros[23] and there are 453 public hospitals in Spain, these savings would enable each hospital to hire more than 2 additional doctors. An increase in the number of obstetricians could help, in turn, to alleviate the need for such long working hours. Importantly, these savings could be materialized without harming neonatal health, given the absence of benefits of these avoidable c-sections.

[20] This figure is calculated with data from 2015, when there were 644 cesareans out of 4027 births in the four hospitals of our sample.

[21] The Spanish National Health System estimated that, for the year 2014, the average cost of a cesarean section without complications was 3,739.06 Euros, while that of a vaginal birth without complications was 2,046.09 Euros. See Ministerio de Sanidad, Servicios Sociales e Igualdad (2014).

[22] The c-section rate for all public hospitals in Spain in 2014 was 22.1%. Assuming that these hospitals have a similar time variation in the c-section rate, removing the excessive c-sections of the early hours of the night would result in a c-section rate of 13.68%. Given that there were 332,252 births, the number of c-sections would decrease from 73,411 to 45,452; that is, a reduction of 27,959 c-sections per year.

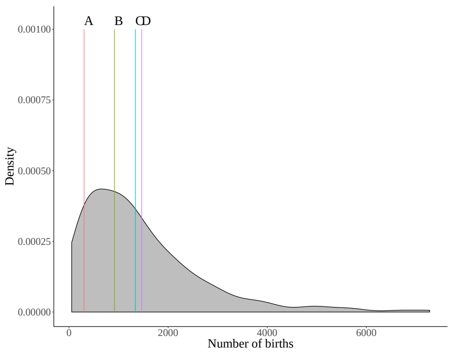
[23] Adecco Healthcare (2017)

1

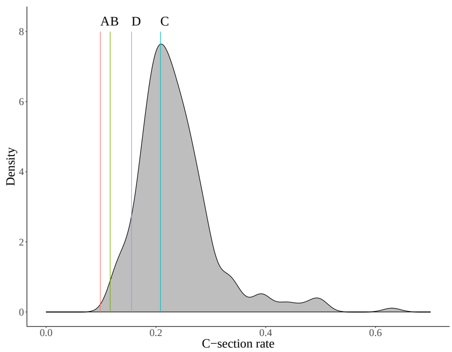
APPENDIX

Appendix 1.A

Figure 1.A.1: Distribution of Number of Births and C-Section Rates in all Spanish Public Hospitals



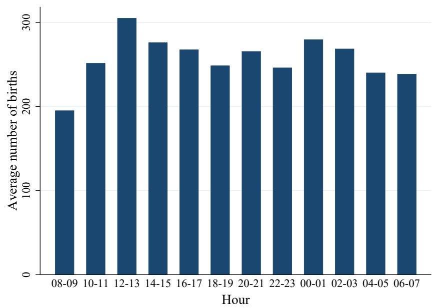
(a) Number of Births Attended in a Year



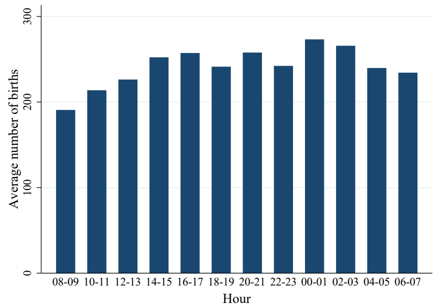
(b) C-Section Rate in a Year

Notes: Figure (a) shows the distribution of the number of births attended in one year for all Spanish Public Hospitals compared to hospitals in our sample (A, B, C and D). Figure (b) shows the distribution of c-section rates in a year for all Spanish Public Hospitals compared to hospitals in our sample (A, B, C and D). Source: our data (2015) and Estadística de Centros Sanitarios de Atención Especializada (2013).

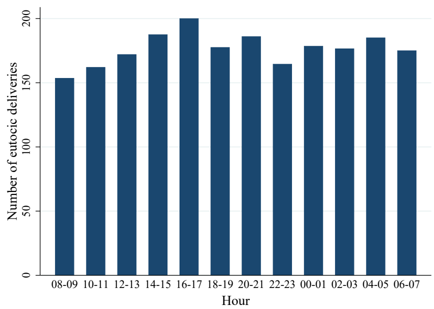
FIGURE 1.A.2: Distribution of Different Types of Births across Times of Day



(a) All Births



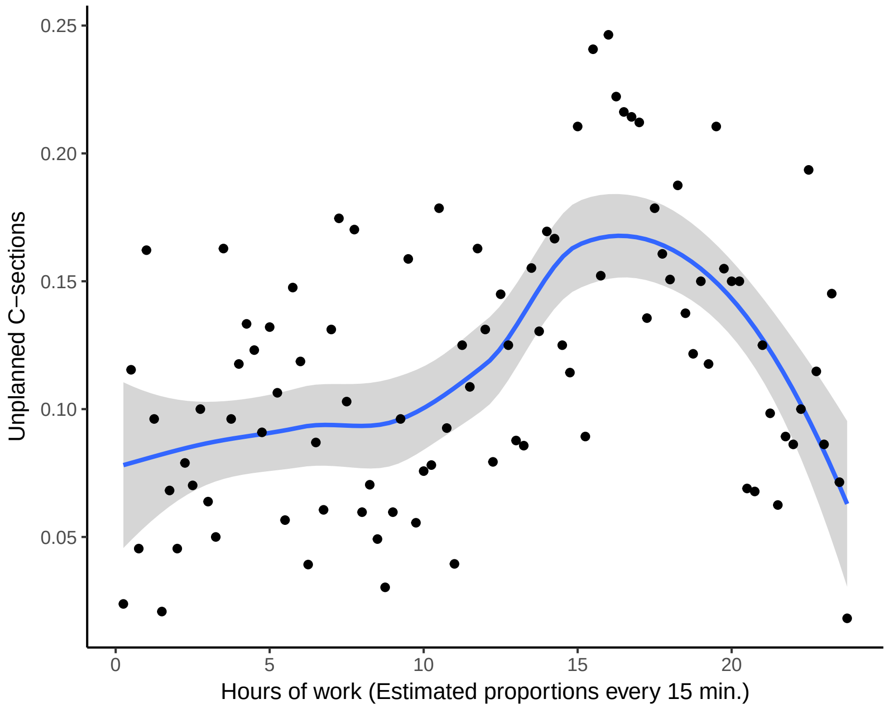
(b) Unplanned C-Sections and Vaginal Births



(c) Eutocic Deliveries

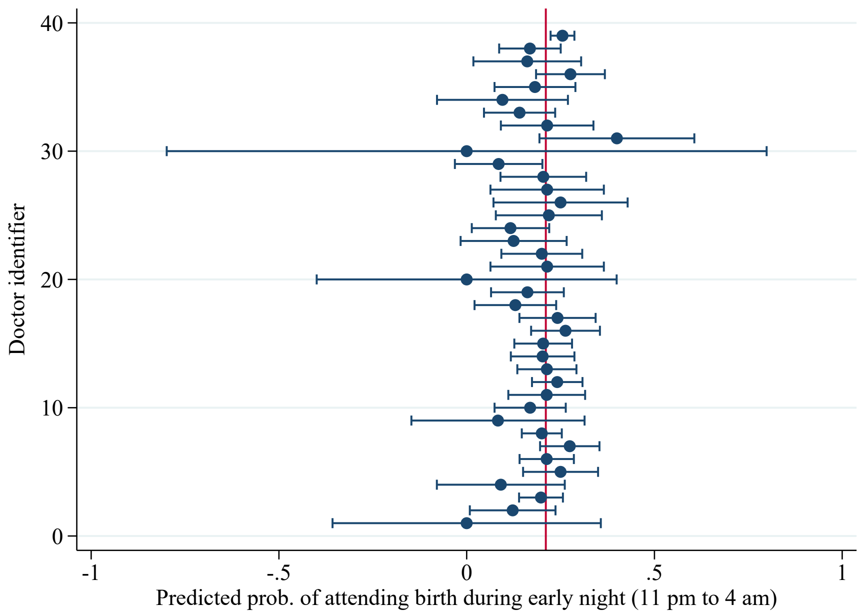
Notes: These figures represent the distribution of different types of births across times of day, grouped by intervals of two hours. Figure (a) represents the number of births per two hours using the full sample of 6,163 observations. Figures (b)-(c) use our usual sample of 5,783 observations. Figure (b) shows the number of births per two hours in this restricted sample, which includes only unplanned c-sections or vaginal births (excluding breech vaginal births), while figure (c) displays the number of eutocic deliveries.

Figure 1.A.3: Proportion of Unplanned C-Sections by Physicians' Hours Worked (Loess Estimate)



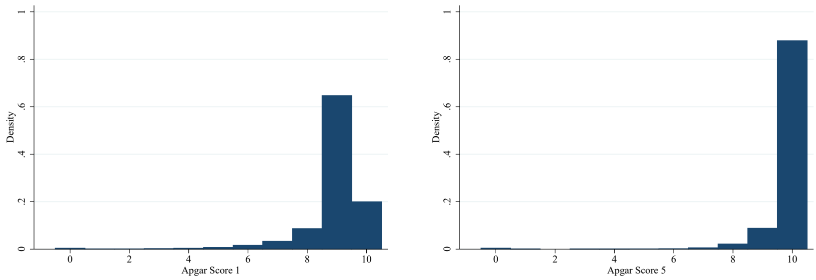
Notes: This figure shows the LOESS or local regression estimate of the proportion of observed unplanned c-sections as a function of a 24h shift, starting at 8 am and finishing at 8 am of the following day with a span of 15 minutes. The shaded area shows the 95% confidence interval.

FIGURE 1.A.4: Predicted Probability by Doctor of Attending Births during the early hours of the night



Notes: The figure shows the probability of attending births during the early hours of the night across different doctors, for a subsample of births for which the doctor identifier was registered (N=3,018). Sample is further restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

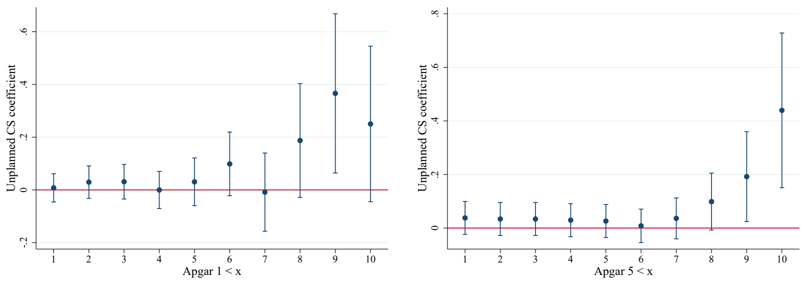
FIGURE 1.A.5: Distribution of Apgar Scores



(a) Apgar Score 1 (b) Apgar Score 5

Notes: These figures show the distribution of Apgar scores for all births. Figure (a) shows the distribution for Apgar scores at minute 1 after birth. Figure (b) shows the distribution for Apgar scores at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

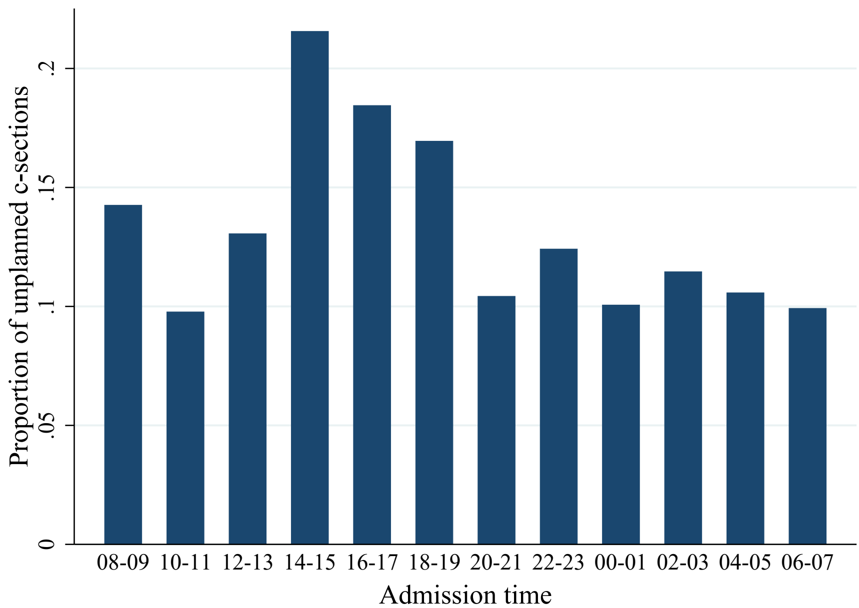
FIGURE 1.A.6: IV Coefficients by Apgar Threshold



(a) Apgar Score 1 (b) Apgar Score 5

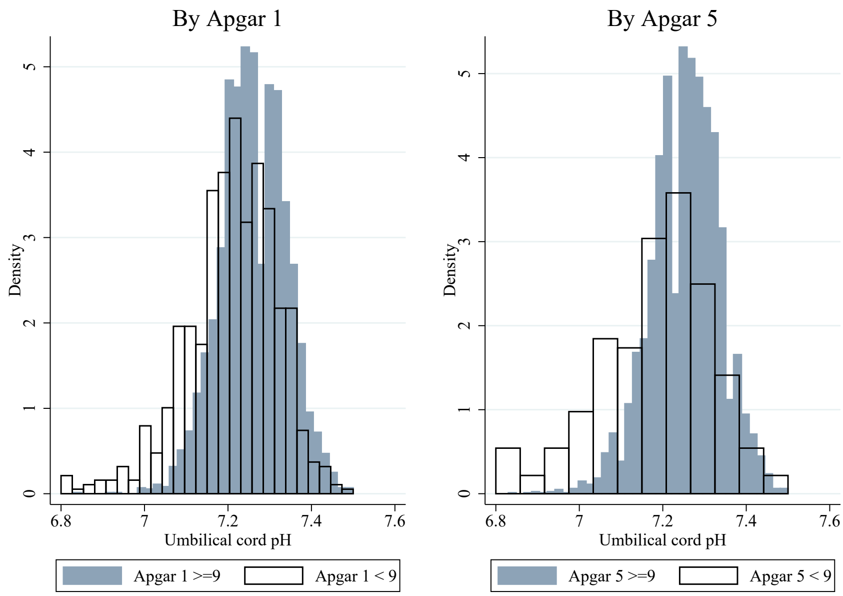
Notes: The figures show the second stage coefficients for the IV regressions of the effect of an unplanned c-section on the probability of Apgar scores being below different thresholds, in regressions with the full set of pregnancy and maternal controls. Figure (a) shows the coefficients for Apgar score at minute 1 after birth. Figure (b) shows the coefficients for Apgar score at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

FIGURE 1.A.7: Proportion of Unplanned C-Sections by Time of Admission



Notes: The figure shows the proportion of unplanned c-sections over the sample of unplanned c-sections and vaginal births, by time of admission to the hospital. Sample is restricted to one hospital (C), single births, unscheduled c-sections and vaginal births (excluding breech babies).

Figure 1.A.8: Distribution of Umbilical Cord pH by Levels of Apgar 1 and 5



Notes: These figures show the distribution of values of umbilical cord pH by Apgar scores above or below 9. Figure (a) shows the distribution for Apgar scores at minute 1 after birth. Figure (b) shows the distribution for Apgar scores at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies)

Table 1.A.1: Summary Statistics

	Mean	SD
A. Mother characteristics		
Mother's age	31.890	5.414
Level of education		
No school	0.032	0.175
Primary school	0.257	0.437
Secondary school	0.523	0.500
University education	0.188	0.391
Non-Spanish	0.250	0.433
Single	0.017	0.130
Mother's weight	65.715	14.536
Mother's height	1.638	2.087
B. Pregnancy characteristics		
Tobacco during pregnancy	0.122	0.327
Alcohol during pregnancy	0.004	0.062
Previous c-section	0.113	0.317
Gestation weeks	39.204	1.785
Multiple pregnancy	0.004	0.064
Obstetric Risk	0.406	0.491
Induction	0.227	0.419
C. Type of birth		
Planned c-section	0.053	0.224
Unplanned c-section	0.112	0.316
Spatula	0.007	0.084
Eutocic	0.687	0.464
Forceps	0.0141	0.118
Breech Vaginal	0.001	0.036
Vacuum	0.125	0.331
D. Newborn outcomes		
Apgar 1	8.884	1.117
Apgar 5	9.793	0.818
Birth weight (in gr.)	3267.970	519.988
Low birth weight (<2500 gr.)	0.068	0.252
Intensive care unit	0.064	0.244
Reanimation	0.084	0.277
Neonatal death	0.004	0.061
Umbilical cord pH	7.254	0.086
Intrapartum pH	7.273	0.073
Male	0.521	0.500
Observations	6163	

Notes: The table shows means and standard deviations for the outcome variables and a set of background variables for all births in our sample of public hospitals.

Table 1.A.2: Summary Statistics of All Births in Spanish Hospitals (2014-2015)

	Mean	SD
Mother's age	32.274	5.449
Non-Spanish	0.180	0.384
Gestation weeks	39.024	1.919
Multiple pregnancy	0.023	0.149
Birth weight (in gr.)	3227.344	531.320
Low birth weight (<2500 gr.)	0.069	0.253
Male	0.516	0.500
Observations	827,692	

Notes: This table shows descriptive statistics from all births in Spanish hospitals in 2014 and 2015. Source: Spanish National Statistics Institute, births microdata.

Table 1.A.3: First Stage: Busy vs. Non-Busy Nights

	(1) Single-birth nights	(2) Multiple-birth nights
Early Night	0.092*** (0.023)	0.054*** (0.012)
Observations	1471	3733

Notes: The table shows the results of the first stage estimation on two different samples: single and multiple birth nights. The coefficients are OLS estimates of the regression of an indicator for an unplanned cesarean birth on an indicator for births during the early hours of the night (from 11 pm to 4 am). Single-birth nights are defined as days in which there is only one delivery from 8 pm to 8 am, whereas multiple-birth nights are those in which more than one delivery occurs during these times. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. The sample is in all cases restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.4: Maternal Characteristics by Type of Birth

	Means		p-value
	Vaginal birth	Unplanned CS	for difference
A. Personal characteristics			
Mother's age	31.622	32.828	0.000
Level of education			
No school	0.033	0.022	0.126
Primary school	0.263	0.206	0.001
Secondary school	0.514	0.609	0.000
University education	0.191	0.164	0.083
Non-Spanish	0.255	0.199	0.001
Single	0.017	0.015	0.662
Mother's weight	65.312	67.830	0.000
Mother's height	1.646	1.595	0.559
B. Pregnancy characteristics			
Tobacco during pregnancy	0.120	0.134	0.277
Alcohol during pregnancy	0.003	0.007	0.089
Gestation weeks	39.320	38.863	0.000
Previous c-section	0.076	0.223	0.000
Obstetric risk	0.367	0.580	0.000
Intrapartum pH	7.288	7.245	0.000
Birth weight	3288.492	3181.038	0.000
Induction	0.214	0.431	0.000
Observations	5098	685	5783

Notes: The table shows means for a set of maternal and pregnancy characteristics by type of birth and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech babies).

Table 1.A.5: Doctor Characteristics by Time of Day

	Means		p-value
	Not early night	Early night	for difference
Male doctor	0.205	0.217	0.538
Number of doctors	1.568	1.603	0.286
Observations	1827	511	2338

Notes: The table shows the mean proportion of male doctors and number of doctors by time of day and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

Table 1.A.6: IV Estimation – Apgar Scores: Standard Errors Robustness

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Unplanned CS	-0.992*	-0.992*	-0.992*	-0.936**	-0.936**	-0.936**
	(0.577)	(0.572)	(0.568)	(0.461)	(0.464)	(0.465)
Mean of Y		8.895			9.798	
Observations		5783			5781	
Cluster (shift)	✓			✓		
Cluster (hospital-shift)		✓			✓	
Robust			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, comparing alternative standard error estimations. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). The first column for each outcome has clustered standard errors at the shift level; in the second column standard errors are clustered at the hospital-shift level, as in our main specification, and in the third column we estimate heteroscedasticity-robust standard errors. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.7: IV Estimation – Full Regression Output Second Stage

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Unplanned CS	-1.122**	-1.147**	-0.992*	-0.956**	-0.987**	-0.936**
	(0.497)	(0.501)	(0.572)	(0.404)	(0.408)	(0.464)
Hospital B	0.188***	0.196***	0.179***	-0.080*	-0.059	-0.071*
	(0.048)	(0.049)	(0.048)	(0.041)	(0.040)	(0.039)
Hospital C	0.234***	0.262***	0.252***	0.170***	0.170***	0.178***
	(0.053)	(0.053)	(0.059)	(0.047)	(0.045)	(0.051)
Hospital D	0.481***	0.503***	0.483***	0.113**	0.152***	0.146***
	(0.058)	(0.060)	(0.062)	(0.046)	(0.047)	(0.048)
Tuesday	-0.005	-0.001	-0.003	-0.021	-0.019	-0.026
	(0.058)	(0.057)	(0.056)	(0.048)	(0.048)	(0.047)
Wednesday	0.043	0.046	0.047	0.063*	0.065*	0.062*
	(0.051)	(0.051)	(0.051)	(0.037)	(0.037)	(0.037)
Thursday	-0.026	-0.023	-0.022	-0.016	-0.015	-0.019
	(0.058)	(0.057)	(0.056)	(0.046)	(0.046)	(0.045)
Friday	0.089*	0.091*	0.093*	0.068*	0.071*	0.068*
	(0.054)	(0.054)	(0.053)	(0.039)	(0.040)	(0.039)
Saturday	0.052	0.056	0.057	0.047	0.050	0.045
	(0.055)	(0.055)	(0.054)	(0.043)	(0.043)	(0.042)
Sunday	0.004	0.007	0.010	0.005	0.007	0.009
	(0.055)	(0.056)	(0.055)	(0.045)	(0.045)	(0.044)
No studies		-0.105	-0.104		-0.088	-0.086
		(0.107)	(0.104)		(0.094)	(0.092)
Secondary school		-0.007	-0.010		0.070*	0.072**
		(0.046)	(0.046)		(0.037)	(0.036)
University education		0.060	0.052		0.089**	0.086**
		(0.048)	(0.047)		(0.038)	(0.037)
Non Spanish		0.057	0.060		0.012	0.015
		(0.042)	(0.042)		(0.032)	(0.032)
Mother weight		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Mother height		0.002	0.002		0.002	0.002
		(0.002)	(0.002)		(0.001)	(0.001)
Mother age		0.000	0.001		0.001	0.001
		(0.003)	(0.003)		(0.003)	(0.002)
Single		-0.170	-0.165		-0.189	-0.179
		(0.144)	(0.142)		(0.122)	(0.120)
Previous c-section			-0.021			0.052
			(0.105)			(0.083)
Prenatal Care 2T			-0.023			0.017
			(0.078)			(0.061)
Prenatal Care 3T			-0.024			-0.015
			(0.170)			(0.097)
Obstetric risk			-0.019			0.032
			(0.037)			(0.031)
Preterm			-0.468***			-0.452***
			(0.138)			(0.131)
Induction			-0.104			-0.014
			(0.076)			(0.063)
Constant	8.783***	8.699***	8.748***	9.831***	9.660***	9.696***
	(0.065)	(0.152)	(0.150)	(0.053)	(0.129)	(0.126)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓
Mean of Y		8.895			9.798	

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). The omitted category for the hospital indicators is Hospital A; for weekdays, it is Monday; for levels of education it is primary school, and for trimester in which prenatal care began it is the first trimester. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.8: IV Estimation – Full Regression Output First Stage

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Early Night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Hospital B	-0.011	-0.006	0.006	-0.011	-0.006	0.006
	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.011)
Hospital C	0.064***	0.056***	0.065***	0.064***	0.056***	0.065***
	(0.012)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)
Hospital D	0.024*	0.036**	0.044***	0.024*	0.036**	0.044***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Tuesday	0.010	0.009	0.004	0.010	0.009	0.004
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Wednesday	-0.010	-0.011	-0.013	-0.010	-0.011	-0.013
	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)
Thursday	-0.004	-0.004	-0.005	-0.004	-0.004	-0.005
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Friday	-0.002	-0.002	-0.007	-0.002	-0.002	-0.007
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Saturday	0.015	0.014	0.007	0.015	0.014	0.007
	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)
Sunday	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009
	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)
No studies		-0.012	-0.005		-0.012	-0.005
		(0.022)	(0.022)		(0.022)	(0.022)
Secondary school		0.026**	0.029***		0.026**	0.029***
		(0.011)	(0.010)		(0.011)	(0.010)
University education		-0.002	0.007		-0.002	0.007
		(0.013)	(0.013)		(0.013)	(0.013)
Non Spanish		0.008	0.007		0.008	0.007
		(0.011)	(0.011)		(0.011)	(0.011)
Mother weight		0.001***	0.001***		0.001***	0.001***
		(0.000)	(0.000)		(0.000)	(0.000)
Mother height		-0.001*	-0.001***		-0.001*	-0.001***
		(0.001)	(0.000)		(0.001)	(0.000)
Mother age		0.004***	0.002**		0.004***	0.002**
		(0.001)	(0.001)		(0.001)	(0.001)
Single		0.003	0.014		0.003	0.014
		(0.031)	(0.032)		(0.031)	(0.032)
Previous c-section			0.151***			0.151***
			(0.021)			(0.021)
Prenatal Care 2T			0.001			0.001
			(0.021)			(0.021)
Prenatal Care 3T			-0.044			-0.044
			(0.028)			(0.028)
Obstetric risk			0.031***			0.031***
			(0.009)			(0.009)
Preterm			0.143***			0.143***
			(0.026)			(0.026)
Induction			0.110***			0.110***
			(0.012)			(0.012)
Constant	0.077***	-0.127***	-0.121***	0.076***	-0.127***	-0.121***
	(0.014)	(0.034)	(0.033)	(0.014)	(0.034)	(0.033)
Observations	5783	5783	5783	5781	5781	5781

Notes: The table shows the first stage coefficients of the IV regression of the effect of an unplanned cesarean birth on Apgar scores 1 and 5. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). The omitted category for the hospital indicators is Hospital A; for weekdays, it is Monday; for levels of education it is primary school, and for trimester in which prenatal care began it is the first trimester. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.9: IV Estimation – Apgar Score < 10

	Apgar Score 1 <10			Apgar Score 5 <10		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	0.283*	0.285*	0.250	0.433***	0.445***	0.439***
	(0.157)	(0.158)	(0.182)	(0.146)	(0.147)	(0.170)
Mean of Y		0.801			0.122	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on the probability of Apgar scores 1 and 5, respectively, being lower than 10. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.10: IV Estimation – Apgar Score < 9

	Apgar Score 1 <9			Apgar Score 5 <9		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	0.380**	0.391**	0.366**	0.189**	0.192**	0.192*
	(0.158)	(0.159)	(0.183)	(0.088)	(0.089)	(0.103)
Mean of Y		0.154			0.034	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on the probability of Apgar scores 1 and 5, respectively, being lower than 9. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.11: Robustness Check: Fetal Distress and C-Sections

	Unplanned CS		Predicted CS	
	(1)	(2)	(1)	(2)
Intrapartum pH	-1.768***		0.018	
	(0.281)		(0.019)	
Intra. pH < 7.2		0.312***		-0.002
		(0.060)		(0.004)
Observations	425	425	425	425

Notes: The table shows the results of OLS regressions of all unplanned cesarean sections and the time-predicted c-sections on indicators of fetal distress. In the first two columns the dependent variable is an indicator equal to one for all unplanned c-sections, while in the last two columns the dependent variable takes the fitted values from the first-stage regression. In the first column for each outcome the explanatory variable is the level of intrapartum of fetal scalp pH, while in the second column is an indicator equal to one if the intrapartum pH is below 7.20. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) for which we have information about the intrapartum pH. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.12: IV Estimation – Apgar Scores: Comparing C-Sections with Eutocic Births

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.179***	-1.218***	-1.161**	-0.907**	-0.954**	-0.942**
	(0.448)	(0.459)	(0.514)	(0.372)	(0.382)	(0.426)
Mean of Y		8.945			9.809	
Panel B. First stage						
Early night	0.090***	0.088***	0.078***	0.090***	0.088***	0.078***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	4886	4886	4886	4884	4884	4884
First-stage F	45.329	43.974	39.192	45.222	43.852	39.102
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, compared to an eutocic birth (a vaginal birth without any instrumentation). The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and eutocic vaginal deliveries. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2

THE LONG-RUN EFFECTS OF CESAREAN SECTIONS

Joint with Mika Kortelainen, Ana Rodríguez-González and Lauri Sääksvuori

2.1 INTRODUCTION

There is little doubt that prenatal health and early childhood circumstances can have long-term effects on mortality, morbidity and human capital development. The theory of the developmental origins of adult health and disease has proven to describe a surprisingly general phenomenon. The effects of prenatal health conditions and early-life events extend to a wide spectrum of educational, cognitive, behavioral and demographic outcomes (Almond et al., 2018).

In human development, the transition from fetal to newborn life at birth is an abrupt event that represents major physiological challenges for the neonates. There is accumulating evidence that many medical and operative interventions at birth are associated with long-term health. Most notably, cesarean delivery for low-risk pregnancies is associated with a wide variety of adverse short- and long-term health outcomes. However, the causal nature of these relationships has received little attention.

The most prominent mechanism thought to mediate the long-term effects of cesarean sections on health and disease emphasizes the importance of early exposure to a diverse range of microbes that adjust the human immune system to appropriately react to extrauterine environment. This general class of mechanisms is often dubbed either as the hygiene hypothesis (Strachan, 1989) or the old friends hypothesis (Scudellari, 2017). According to these hypotheses, children born by cesarean section lack the beneficial exposure to their mother's vaginal microbiome and are more prone to develop immune-mediated diseases.

Cesarean section is the most commonly performed major surgery in many countries. Understanding the consequences of cesarean sections on later-life health and human capital development is important from a number of perspectives varying from clinical decision making to economic and health policy. The rapidly growing incidence of cesarean sections across the globe suggests that even small increases in mortality and morbidity due to C-sections would lead to large reductions in life expectancy and substantial losses of human welfare.[1]

This paper provides new evidence on the effect of potentially avoidable cesarean sections on several relevant health outcomes. To identify the causal effect and abstract from cases where C-sections respond to a clear medical indication, we exploit variation in physician demand for leisure. We show that the probability of unscheduled C-section increases substantially during the normal working hours (8am – 4pm) on working days that precede a leisure day. Importantly, we find that

[1]Cesarean section rates have increased in the US from 20.7 percent in 1996 to 32.9 percent in 2009 (Currie and Macleod, 2017). In OECD countries, the rate of cesarean sections has increased from 20 percent in 2000 to 25 percent in 2013 (OECD, 2013). Currently, the highest rates of cesarean sections are reported in many of the world's most populous countries including among others China (41.3 percent in 2016) and Brazil (55.6 percent in 2015). Boerma et al. (2018) review the disparities in C-section use around the world.

these excess C-sections are not driven either by selection of different mothers giving birth at these times or by advancing births that would have been cesarean deliveries in any event.

Using fine grained data on birth times and intrapartum diagnoses, we show that the increased likelihood of cesarean sections during the normal working hours on days that precede a leisure day is coupled with the increased use of more discretionary diagnoses. Moreover, we observe that physician demand for leisure does not affect mothers who are in the medical profession. Our data lend substantial support for the contention that the excess numbers of unplanned cesarean deliveries observed during the normal working hours on days that precede a leisure day are largely driven by physician incentives. We use this time variation as an instrument for C-section. We provide a detailed discussion and numerous robustness checks to support the validity of the required identification assumptions.

We investigate the effects of cesarean sections on infant and children outcomes using a comprehensive and precise administrative data resource which includes birth and health records for all children born in Finland between 1990 and 2014. We follow entire birth cohorts from birth to teenage years and use detailed diagnosis data to study the causal effects of cesarean sections on children's health. We focus on outcomes whose onset is hypothesized to be influenced by cesarean delivery: asthma and other atopic diseases, type 1 diabetes and obesity. These are among the most common chronic conditions in childhood (Torpy, 2010).[2]

[2]Understanding and quantifying the potential contribution of C-sections to the development of these diseases is not limited to medical practice and health policy. Chronic health conditions cause an immense financial burden to households and public health care financing. The total cost of asthma in the working age population was estimated to be $24.7 billion during 1999-2002 in Europe (Global Asthma Network, 2018). The two other atopic diseases we investigate imply high costs: atopic dermatitis has been estimated to cost at least $5.3 billion (in 2015 USD) in the US (Drucker et al., 2017). The estimated annual cost of allergic rhinitis is in the range of $2–5 billion (in 2003 USD) (Reed et al., 2004). Type 1 diabetes has been found to cost $14.4 billion a

Our instrumental variable estimates suggest that avoidable C-sections increase the probability of asthma diagnosis from early childhood onward. This effect is clinically and economically relevant. However, we do not find consistent evidence that cesarean sections affect the probability of developing atopic diseases at large, type 1 diabetes or obesity.

We complement our instrumental variables estimates using a differences-in-differences model with family fixed effects that compares the health gap between siblings in families where the second child was born by unplanned C-section with the health gap between siblings who were born by vaginal delivery. The results from our supplementary empirical strategy support our main findings. These estimates suggest that unplanned C-sections increase the risk of childhood asthma and enable to rule out meaningful effects on other atopic diseases, type 1 diabetes and obesity. We provide several sensitivity checks that suggest that the effect on asthma is unlikely to be explained by negative selection.

Our results are consistent with the hypothesis that the mode of delivery may influence the development of the immune system and have long-term effects on health and disease. However, our results paint a more nuanced picture about the long-term effects of cesarean deliveries than existing evidence based mostly on associations. Our findings suggest that C-sections cause a much narrower spectrum of diseases than currently hypothesized and call for a careful analysis on the relationships between the delivery mode and long-term health.

Our paper relates to an important literature estimating the effects of early interventions on long-term health and human capital development. Moreover, we contribute at least in three ways to a nascent economics literature on the effects of treatment choices at birth. First, we investigate the long-term effects of unplanned C-sections on chil-

year in medical costs and lost income in the US (Tao et al., 2010). Finally, childhood obesity, which has been on the rise in recent years, has been calculated to imply $19,000 per child in lifetime medical costs in the US (Finkelstein et al., 2014).

dren. To evaluate the costs and benefits of C-sections, it is crucial to investigate long-term effects, as potential alterations of the immune system and long-run consequences of C-sections are not necessarily visible at birth and in early childhood. Moreover, we report age-by-age estimates for entire cohorts from birth to teenage years and provide evidence about the effects of early life events during the middle childhood, thus expanding our knowledge about the "missing middle" years.[3] Existing papers investigating the effects of potentially avoidable C-sections have concentrated on neonatal outcomes or short-term effects.[4] Costa-Ramón et al. (2018) investigate the effects of cesarean sections on neonatal health using time variation in unplanned C-section rates. Card et al. (2019) study the short-term health effects of hospital delivery practices using relative distance from a mother's home to hospitals with high and low C-sections rates.[5]

Second, we study the effects of discretionary unplanned C-sections that could potentially be avoided, while existing papers have not been able to separate planned (elective) and unplanned C-sections or have concentrated on C-sections with a clear medical indication. Hannah et al. (2000), Jensen and Wüst (2015) and Mühlrad (2017) show that

[3] Almond et al. (2018) discuss that, due to data availability, most of the literature analyzes the effect of early life events on birth or adult outcomes. This implies that we have little knowledge about how developmental trajectories are affected by policies or shocks experienced over the life course. They refer to this gap in the literature as the "missing middle".

[4] To our knowledge, the only paper looking at longer-term effects is by Jachetta (2015), who explores the relation of cesarean delivery with hospitalizations using regional variation in medical malpractice insurance premia in the US as an instrument for C-sections. However, the instrument used in that paper does not necessarily allow for credible causal inference, since the author finds that higher premia also predict delayed prenatal care, lower birth weight and reduced gestational age.

[5] A few papers have also examined the effects of cesarean sections on mothers. Halla et al. (2016) study the effects of C-sections on fertility and maternal labor supply. Tonei (2019) studies the impact on mental health for mothers with breech babies who undergo a C-section. Our findings on children health complement these maternal results and contribute to obtaining a more complete picture of the effect of cesarean sections.

breech babies can benefit from C-section delivery. However, these results concern medically necessary C-sections in a specific high-risk group and do not readily generalize to cesarean deliveries in general or for avoidable unplanned C-sections, in particular. While C-sections are often life-saving at the top of the risk distribution (Currie and Macleod, 2017), more evidence is required about the effects of discretionary C-sections that could be potentially avoidable.

Third, to evaluate causal effects of C-sections, we use two different identification strategies based on somewhat different assumptions. Our instrumental variable strategy builds on previous work using time variation in C-section rates in combination with high-quality administrative data. Moreover, we employ a differences-in-differences research design that has not been used in previous papers on C-sections. In addition, for both methods we provide several pieces of evidence that support the credibility of the identification assumptions. Thus, by using two different strategies, we hope to provide more reliable evidence on the causal effects of avoidable unscheduled interventions at birth on children both in the short and long run.

The paper is structured as follows. Section 2.2 provides background information about the biological mechanisms hypothesized to mediate the effects of mode of delivery on infant outcomes, about the different types of cesarean sections, and about the institutional context of our analysis. Section 2.3 introduces the data, provides key descriptive statistics and lays out our econometric approach. Section 2.4 reports our main results. Section 2.5 presents robustness checks and additional evidence to support our main conclusions. The last section concludes.

2.2 BACKGROUND

2.2.1 *Mechanisms*

A large body of literature documents the developmental origins of health and disease. The process of labor can be seen as one crucial step in adaptation to the extrauterine environment. The prevailing evidence highlights the role of vaginal delivery as an important early programming event with potentially life-long consequences (Hyde et al., 2012). While there is strong consensus that medically indicated cesarean sections decrease the risk of fetal death at birth, the absence or modification of vaginal delivery has been linked to several adverse health outcomes and anomalies in human development. In the following, we summarize some of the most widely acknowledged findings to understand how C-sections might have long-lasting effects on health and human development.

It is well-recognized that early exposure to microbes is necessary to train the human immune system to react appropriately to environmental stimulation. The original formulation of the theory, dubbed as the hygiene hypothesis, states that the lack of early childhood exposure to infectious agents and symbiotic microbes increases susceptibility to multiple autoimmune diseases by suppressing the natural development of the immune system (Strachan, 1989). Lately, refinements to the original formulation, known as the old friends hypothesis, have challenged the role of infectious pathogens and highlight the importance of early exposure to a diverse range of harmless microbes to strengthen the human immune system and combat the threat of environmental pathogens (Scudellari, 2017).

Mode of delivery may affect early exposure to microbes through several channels. First, bacteria from the mother and the surrounding environment colonize the infant's gut during birth (Neu and Rushing, 2011). Exposure to the maternal vaginal microbiota is interrupted in a cesarean birth and externally derived environmental bacteria play an important role for the infants' intestinal colonization. Consequently, infants delivered by C-sections acquire a microbiota that differs from that of vaginally delivered infants (Dominguez-Bello et al., 2016). Second, the transfer of microbiota continues through breastfeeding after birth. Breast milk contains a number of bioactive components that can have an important impact on infant's microbiota composition and health (Collado et al., 2015). The negative association between cesarean sections and the initiation of breastfeeding provides an additional mechanism to explain the differences in microbiota by type of birth (Prior et al., 2012).

The potential biological mechanisms are consistent with the reported associations between cesarean delivery and adverse infant outcomes. These studies relate cesarean deliveries to a marked increase in the susceptibility of multiple immune and metabolic conditions. Even though cesarean deliveries have been associated with a broad array of immune-mediated diseases, recent meta-analyses conclude that C-sections are most robustly related to asthma, atopic diseases, type 1 diabetes and obesity (Blustein and Liu, 2015; Keag et al., 2018; Cardwell et al., 2008; Thavagnanam et al., 2008; Peters et al., 2018; Bager et al., 2008).[6] However, the causal nature and clinical relevance of these relationships remains largely unknown.[7]

[6]In addition to health outcomes, literature has associated cesarean sections with worse cognitive and emotional development (Bentley et al., 2016).

[7]Hyde et al. (2012) summarize evidence from 14 RCTs that compare the effects of cesarean and vaginal deliveries on infant health. All these studies are small RCTs conducted in populations of at risk babies (e.g. breech delivery). These studies have had exceptionally large problems to achieve target recruitment and do not include long-term follow-ups. Overall, there exist no RCTs to date that would enable to investigate the long-term effects of cesarean sections on infant health. Hyde and Modi (2012) report evidence from survey studies that investigate the perceived acceptability of randomizing the mode of delivery to address long-term health outcomes in low-risk pregnancies. The perceived acceptability of randomizing the mode of delivery in healthy, term, cephalic and singleton pregnancies remains low among obstetricians and mothers, suggesting that adequately powered large-scale RCTs to compare the effects of cesarean and vaginal deliveries on long-term outcomes may remain unrealized in the near future.

2.2.2 *Classification of Cesarean Sections*

Cesarean sections are performed for several indications at different stages of the pregnancy. Cesarean sections are classified either as scheduled (elective) or unscheduled operations. Scheduled C-sections occur without attempted labor and are agreed upon in advance. The large majority of scheduled C-sections are performed during the regular working hours (8am — 4pm) from Monday to Friday. Medical indications that make scheduled C-sections advisable include, among others, multiple pregnancies with non-cephalic presentation of the first fetus or placenta previa. We exclude all scheduled C-sections from our sample.

Most C-sections are performed with no scheduled intervention after spontaneous or medically induced onset of labor. Unscheduled C-sections are surgeries where an attempt of vaginal birth is transformed to a cesarean delivery after the mother has been admitted to a hospital. Unscheduled C-sections are classified by urgency. Emergency C-sections are performed within 30 minutes of the decision, due to an immediate threat to the life of the mother or the baby (NICE, 2011). However, most unscheduled C-sections are performed without such immediate threat. The optimal timing and indication for these operations are imprecise and give large discretion to the clinician. Slow progression of labor or cephalopelvic disproportion are examples of diagnoses that may require an unplanned non-urgent cesarean section. There is wide variation among clinicians in the use of discretionary diagnoses that justify C-sections (Barber et al., 2011; Fraser et al., 1987). Our data contains the registered diagnosis linked to the C-section for a subsample of births. These observations enable us to verify that the peaks in unplanned C-sections are coupled with the use of more discretionary diagnoses.

2.2.3 *Institutional Context*

Finland has universal public health coverage. Comprehensive pre- and postnatal care services are included in the publicly provided services. There are no private medical institutions running maternity wards. Consequently, all deliveries take place in public hospitals. All medical expenses related to prenatal care, delivery and postnatal care are fully covered by the public health care system.

Pregnant women usually give birth in the nearest hospital. Only high-risk pregnancies are systematically directed to a higher-level hospital for obstetric care and delivery. Expectant women do not have pre-assigned midwifes or physicians for the delivery. Midwives take care of the delivery in all hospitals, while physicians have the ultimate responsibility for obstetric care, decide on the type of delivery and perform C-sections. There are no delivery units led by midwifes. The C-section rate (15.5% in 2015) is relatively low from an international perspective (OECD, 2017).

The regular working shifts for physicians are from 8 am to 4 pm from Monday to Friday. The on-call hours for physicians may not exceed 24 hours during the regular working week and last typically from 8 am to 8 am. On weekends, the on-call hours for physicians are from 8 am to 9 am on next day.[8] Midwives follow the same rotation regardless of the type of day and work in three shifts of around 8 hours.[9]

[8] Even though the statutes that govern on-call arrangements have changed in recent years, during most years covered in our data, small hospitals with less than 1000 annual births could autonomously decide their on call arrangements. In certain hospitals, physicians were allowed to be at home while on duty, if they could arrive to the hospital within 30 minutes from home.

[9] An example of midwives' schedules: (i) from 7 am to 3 pm, (ii) from 2 pm to 9.30 pm, and (iii) from 9.15 pm to 7.15 am.

2.3 DATA AND METHODS

2.3.1 *Data*

The two main data sources used in our analysis are the Finnish Medical Birth Register and the Hospital Discharge Register. The Finnish Medical Birth Register was established in 1987. This administrative data resource includes data on all live births and on stillbirths of fetuses with a birth weight of at least 500 grams or with a gestational age of at least 22 weeks. The register includes information on maternal background, health care utilization, and medical interventions during pregnancy and delivery. It also includes mother's diagnoses during delivery (ICD-10 codes) and newborn outcomes until the age of 7 days. From 1990, the register contains detailed information about the type of C-section (scheduled vs. unscheduled). These data are collected at all delivery hospitals.

We exclude from our sample planned C-sections and multiple pregnancies. For our instrumental variable strategy, we focus only on first births.[10] Our analysis sample includes 392,560 deliveries that took place from 1990 to 2014. For the differences-in-differences analysis, we focus on both first and second births from families where the first child was born by vaginal delivery (more details are provided in section 2.3.2.2). The analysis sample consists of 645,292 children from 322,646 sibling pairs. There are 43 hospitals in our sample. Table 2.A.1 shows summary statistics for all births in Finland between 1990 and 2014.

We match the Finnish Medical Birth Register to the Finnish Hospital Discharge Register, which contains information about the diagnosed

[10]We follow a common practice in literature and focus on first births, which also allows us to keep just one birth per mother, and abstract from a potential source of correlation between the observations. First-time mothers are also the group of mothers where we find larger variation. Given the faster pace of labor in higher-order births (NICE, 2014) and the high risk of repeated C-section, there is less room for discretion in the decision to perform an unplanned C-section in subsequent deliveries. Our results are qualitatively similar but less precise when we include higher order births.

medical conditions, medical operations, and the date of diagnoses. This hospital register contains all inpatient consultations in Finland from 1990 to 2013. From 1998, the data include all outpatient visits to hospitals. All diagnoses are coded using the International Classification of Diseases (ICD) tool.[11]

We explore two sets of outcome variables. First, to test whether unplanned C-sections have an impact on neonatal health, we analyze indicators of neonatal health included in the birth register. We study Apgar scores one minute after birth, admission to intensive care unit (ICU), need of assisted ventilation and early neonatal mortality (defined as neonatal death in the first week of life).[12] Second, we study longer term outcomes using detailed inpatient and outpatient diagnosis data from the Finnish Hospital Discharge Register. We use primary diagnoses.[13] To maintain a relatively large sample size, we follow individuals from birth until age 15. We focus on the four metabolic and immune-related conditions that have been most robustly associated with cesarean delivery: asthma, atopic diseases (atopic dermatitis and allergic rhinitis), type 1 diabetes and obesity. Table 2.A.2 in the appendix provides more detail about each of these diagnoses.

[11] Diagnoses for years from 1990 to 1995 are recorded using ICD-9 classification. Diagnoses from 1996 onwards are recorded using ICD-10 classification. The quality and completeness of the Finnish Hospital Discharge Register has been assessed in multiple validation studies that have compared recorded data entries with external information. The completeness and accuracy of the data are found to be exceptionally high (Sund, 2012). We assess to what extent our data is able to identify the individuals with a certain diagnosis in the Results section.

[12] Apgar scores result from the examination of the newborn by the midwife or pediatrician one minute after the birth. Five different dimensions are measured and graded from 0 to 2: appearance (skin color), pulse (heart rate), grimace (reflex irritability), activity (muscle tone), and respiration. The resulting score takes values from 1 to 10.

[13] We replicated all our analysis using both primary and secondary diagnoses. All results remain unchanged. Results are available upon request.

2.3.2 *Empirical strategy*

We aim to estimate the impact of a cesarean delivery on child's health at birth and older ages. We define a binary variable CS_i that takes value 1 if the delivery is an unplanned C-section and 0 if it is a vaginal delivery. Thus, we aim to estimate the following equation:

$$Y_i = \beta_0 + \beta_1 CS_i + X_i'\beta_2 + \delta_m + \lambda_y + \phi_h + \varepsilon_i, \qquad (2.1)$$

where Y_i is the health outcome of infant i, X_i is a vector of covariates and δ_m, λ_y, ϕ_h are fixed effects for the month, year, and hospital of birth, respectively.[14]

The estimation of equation (2.1) is, however, likely to provide biased estimates of β_1 due to potential selection into cesarean birth.[15] To study the causal effects of cesarean delivery on health, we exploit two different empirical strategies.

2.3.2.1 IV strategy: Variation by time and type of day

Our instrumental variable strategy exploits the higher likelihood of being born by C-section during the normal working shift on pre-leisure days compared to regular working days. We use the interaction between the type of day and work shift as an instrument for the mode of delivery.

Figure 2.1 presents the predicted probability of unplanned C-section delivery by hour and type of day. We adjust for hospital, month and

[14]The vector of covariates includes the gender of the baby, the mother's marital status, nationality, socioeconomic status, age and smoking status. In addition, we include a wide range of pregnancy and delivery related indicators that include in-vitro fertilization, amniocentesis during pregnancy, ultrasound during pregnancy, gestational diabetes, maternal hospitalization due to hypertension, maternal hospitalization due to placenta previa, maternal hospitalization due to eclampsia, gestational weeks, induced labor, prostaglandin pre-induction, epidural use, and laughing gas anesthesia.

[15]Figure 2.A.1 in the appendix shows that mothers and babies who undergo a C-section are very different from those mothers and babies who undergo a vaginal delivery.

year of birth fixed effects. Figure 2.1a plots the distribution of C-sections over a 24-hour cycle for working days that precede a leisure day compared to other working days.[16] We find that substantially more C-sections are performed during regular working hours on days that precede a leisure day compared to the rest of working days. Figure 2.1b presents the predicted probability of having an unplanned C-section by work shift and type of day. We find that the gap in C-section rates between a day that precede a leisure day and the rest of working days emerges only during the regular working hours (from 8 am to 4 pm).

Importantly, we find that the excess C-sections performed in days that precede a leisure day are not driven by advancing births that would have been cesarean deliveries in any event. We do not observe any relative fall in C-sections during the evening hours preceding a leisure day compared to the evenings of regular working days (Figure 2.1a) or during the leisure day (Figure 2.A.2 in the Appendix).[17] These observations suggest that physicians perform C-sections during the regular working hours on pre-leisure days that would not have been performed otherwise.

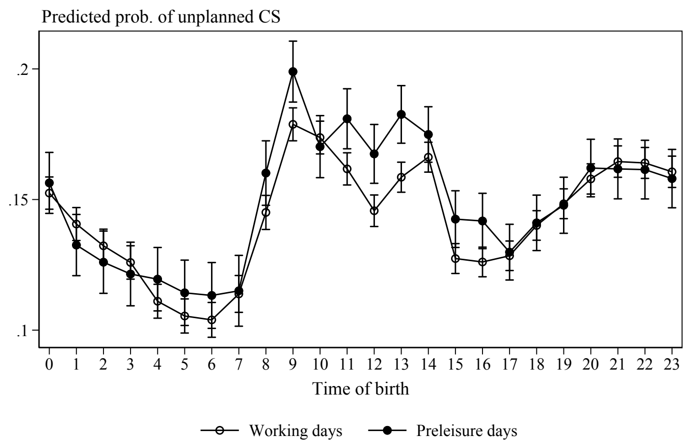
The time pattern of C-sections is consistent with previous work by Brown (1996) and Halla et al. (2016) that documents an increase in C-section rates on days that precede a leisure day. Halla et al. (2016) exploit this variation in an instrumental variable framework to study the impact of delivery mode on maternal fertility and labor supply. Like the existing literature, we attribute the pre-leisure anomaly in the time pattern of C-sections to physician demand for leisure. This incentive arises from the higher time cost and uncertainty of vaginal births. A

[16]Working days that precede a leisure day include Fridays and days preceding public holidays. Table 2.A.3 documents all public holidays in Finland. Friday is not considered a working day that precedes a leisure day if it is a holiday.

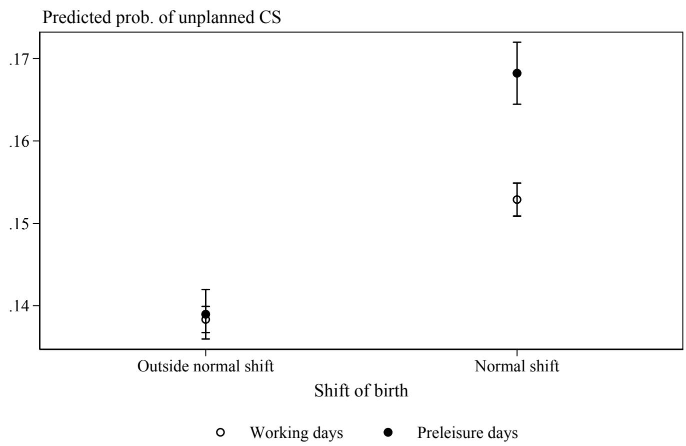
[17]This figure compares the predicted probability of unplanned C-section by hour separately for Saturdays or holidays (the leisure day following the pre-leisure day) and Sundays (a leisure day that is not preceded by a working day). We do not see any relative drop in the C-section rate on Saturdays compared to Sundays at any time of day.

FIGURE 2.1: Predicted probability of unplanned C-section

(a) By time of birth



(b) By shift of birth



Notes: Figure (a) presents the predicted probability of unplanned C-section by hour and type of day. Figure (b) shows the predicted probability of unplanned C-section by shift and type of day. Both figures adjust for hospital, month, and year of birth fixed effects. Pre-leisure days include working days that precede a Finnish public holiday or a weekend, while working days include the rest of working days. Sample is restricted to singleton first births which are either unscheduled C-sections or vaginal births.

cesarean section takes on average 30-75 minutes and is perceived as a relatively easy surgical intervention with low complication rates (NICE, 2011).The average duration of labor for first-time mothers who have a vaginal birth is 11 hours (NICE, 2014).

We provide two pieces of complementary evidence to validate that the excess rate of C-sections is not driven by medical factors. First, we build on previous evidence that some medical diagnoses linked to a cesarean birth are more discretionary than others. Dystocia (prolonged or obstructed labor), one of the most common indications for primary cesarean section, is believed to provide the greatest room for diagnostic discretion (Fraser et al., 1987). The number of dystocia diagnoses has been shown to strongly respond to physician incentives (Evans et al., 1984; Fraser et al., 1987; McCloskey et al., 1992). We examine if there is an excess number of dystocia diagnoses during regular working hours on pre-leisure days. Our results (Table 2.A.4) show that giving birth during the regular hours on a pre-leisure day increases the probability of having a dystocia diagnosis compared to other working days. Importantly, we do not find this temporal pattern for medical emergencies, for which there should not be any room for discretion. In particular, we find that our instrument does not predict additional examinations of the fetus during labor, which doctors should perform if there are any signs of fetal suffering.[18]

Our second piece of evidence builds on the literature showing that physician mothers are less likely to receive C-sections driven by financial incentives (Johnson and Rehavi, 2016). Consequently, we expect that the probability of having a C-section does not respond to physician demand for leisure among physician mothers and other medical professionals. Our results (Table 2.A.5) support this hypothesis. We do not find that medical professionals have an increased risk of having a C-section during the regular shift on pre-leisure days, while we do

[18]We examine whether physicians take measurements of intrapartum or fetal scalp pH, which proxies the oxygen saturation of fetal blood during labor.

find this increase for non-medical mothers with an equivalent level of education.[19]

We exploit the variation in the probability of unplanned C-sections by time and type of day and adopt an instrumental variable approach. We first estimate a standard two-stage least squares (2SLS) with the following first stage:

$$CS_i = \gamma_0 + \gamma_1 NS_i + \gamma_2 Preleisure_i + \gamma_3 NS_i \times Preleisure_i + X_i' \gamma_4 + \delta_m + \lambda_y + \phi_h + \upsilon_i \quad (2.2)$$

and the corresponding second stage:

$$Y_i = \alpha_0 + \alpha_1 NS_i + \alpha_2 Preleisure_i + \alpha_3 \widehat{CS}_i + X_i' \alpha_4 + \delta_m + \lambda_y + \phi_h + \varepsilon_i \quad (2.3)$$

where NS_i is a dummy that takes value 1 for births that take place during the normal shift (from 8 am to 4 pm) and 0 otherwise, $Preleisure_i$ takes value 1 for Fridays or working days preceding a Finnish public holiday and 0 for other working days, $\widehat{CS}_i$ in equation (2.3) are the predicted C-sections from the first stage, X_i is the vector of individual controls,[20] and δ_m, λ_y, ϕ_h are month, year, and hospital of

[19] Our definition of medical professionals includes physicians, midwifes and nurses. Our observation relates to a large literature on physician-induced demand in health care. Since the work of Arrow (1963), it has been recognized that asymmetric incentives between physicians and their patients are a central feature of the medical marketplace. The role of financial incentives on the supply of cesarean sections has been documented by Gruber and Owings (1996). Johnson and Rehavi (2016) observe that financial incentives have a particularly large effect on the probability of having a cesarean section among non-physicians. Our results complement the literature on physician-induced demand and show that the excess rate of C-section on pre-leisure days is restricted to non-medical professionals.

[20] Gender of the baby, mother's marital status, nationality, socioeconomic status, age, smoking status, and the following pregnancy and delivery characteristics: gestational weeks and indicators for in-vitro fertilization, amniocentesis during pregnancy, ultrasound during pregnancy, gestational diabetes, maternal hospitalization due to hypertension, maternal hospitalization due to placenta previa, maternal hospitalization due to eclampsia, induced labor, prostaglandin pre-induction, epidural use, and laughing gas anesthesia.

birth fixed effects, respectively. The interaction between regular working hours and a day preceding a leisure day will serve as an instrument. As a result, we will be comparing mothers who give birth in the same hospital during the same shift, but on different types of days (working days preceding a leisure day or other working days). We expect a positive $\widehat{\gamma}_3$ due to increasing physician demand for leisure on days preceding a weekend or public holiday.

Our instrumental variables estimation needs to meet three conditions to yield valid estimates. First, the instrument should strongly influence the probability of C-section (first stage). Second, there should be no selection of mothers who give birth during the regular shift on different types of days. Finally, being born during the regular shift on pre-leisure days, compared to other working days, should only affect child outcomes through the increased probability of being born by C-section (exclusion restriction).

Table 2.1 shows the results from the estimation of the first stage. Column (1) shows the first stage estimates including month, year, and hospital fixed effects. Column (2) includes a richer set of controls. These estimates show that being born during the normal shift increases the probability of C-sections for all working days. Moreover, being born during the normal shift on pre-leisure days increases the probability of C-section by 1.5 percentage points. The first stage F-statistics are larger than 25 in both specifications. Following the common critical values for weak instruments (Stock and Yogo, 2005), we can reject the null hypothesis that the instrument is weak.

Figure 2.2 shows that our instrument does not predict a large set of maternal and pregnancy characteristics, including medical conditions that could predict a C-section. This indicates that mothers giving birth during the regular shift on pre-leisure days compared to other working days are similar in observable characteristics, suggesting that the observed increase in C-sections at these times cannot be explained by selection.

Table 2.1: First stage

	Unplanned CS	
	(1)	(2)
Normal shift	0.015***	0.017***
	(0.001)	(0.001)
Preleisure day	0.001	-0.002
	(0.002)	(0.002)
Normal shift× Preleisure	0.015***	0.014***
	(0.003)	(0.003)
Observations	392561	392561
Controls	NO	YES
$\overline{Y}$	0.145	0.145
First-stage F	26.650	25.209
Adjusted R^2	0.008	0.070

This table shows estimates from the first stage (see equation (2.2)). All specifications include hospital, year and month of birth fixed effects. Controls: gender, maternal age, marital status, nationality, mother occupation (long-term unemployed, high-skilled white collar, low-skilled white collar, manual worker, student, other), whether mother smoked during pregnancy, high/low number of prenatal visits, IVF, gestation weeks, induced labor, prostaglandin preinduction, epidural or laughing gas anesthesia. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Finally, regarding the exclusion restriction, we focus on births that take place on working days, when hospital resources and quality of care should be constant. Moreover, to compromise our empirical strategy, any change in the quality of care would need to happen on pre-leisure days only during the regular working hours. We provide numerous supplementary analyses in section 2.5.1 that reinforce the credibility of this assumption.

The two-stage least squares estimator enables us to identify a local average treatment effect (LATE). This is the effect of C-sections for infants whose mothers' mode of delivery is sensitive to the subjective assessment of the physician. More accurately, we capture births where the type of day affects the decision of the doctor to perform a C-section during the normal shift. The counterfactual for these births is unlikely to be exclusively a cesarean section later on, given that we do not find a relative drop in C-sections on pre-leisure days after the normal shift or during the following day. The LATE will not be informative of the effect of medically indicated C-sections, as those are not affected by leisure incentives. Moreover, the LATE does not capture the effect of unplanned C-sections for babies who had a very fast delivery, leaving no room for physician discretion.

Our primary health outcomes and the endogenous variable are binary. Consequently, besides the 2SLS models we estimate (recursive) bivariate probit models. These specifications mirror equations (2.2) and (2.3) and assume that cesarean delivery (CS_i) and the binary indicator of health Y_i are determined by the following latent indices:

$$CS_i = 1[\rho_1 NS_i + \rho_2 Preleisure_i + \rho_3 NS_i * Preleisure_i + X_i' \rho_4 + \delta_m + \lambda_y + \phi_h + \nu_i > 0] \tag{2.4}$$

$$Y_i = 1[\pi_1 NS_i + \pi_2 Preleisure_i + \pi_3 CS_i + X_i' \pi_4 + \delta_m + \lambda_y + \phi_h + \xi_i > 0] \tag{2.5}$$

where (ν_i, ξ_i) follow a bivariate standard normal distribution with unknown correlation. These equations can be estimated through maximum likelihood. Identification in this setting relies on the same assumptions that are needed to estimate the 2SLS model together with an additional assumption about the joint normality of the error terms.

Bivariate probit estimation is expected to present substantial advantages in the context of this paper. The bivariate probit estimation is shown to be more efficient and less biased than 2SLS when treatment

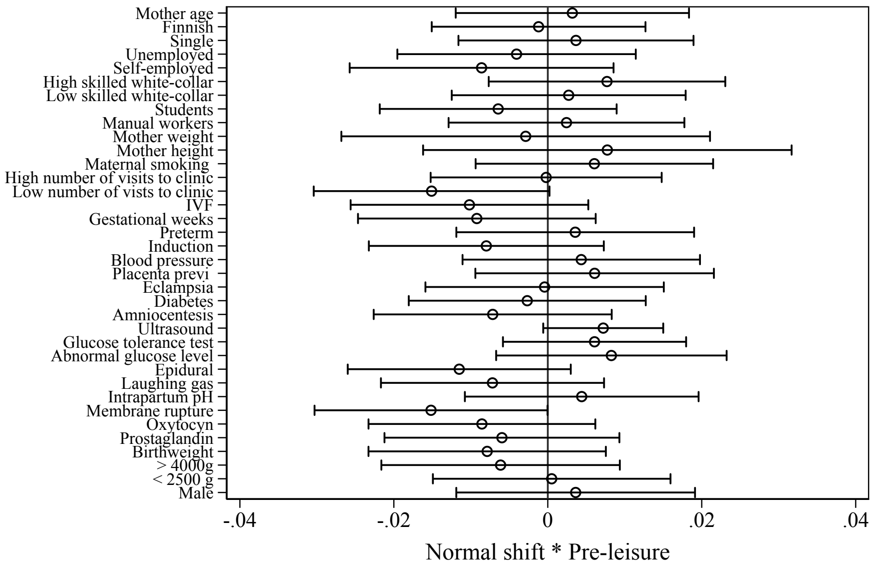


FIGURE 2.2: Instrument and baseline characteristics

Notes: The figure represents the coefficients and 95% CI from separate regressions of each (standardized) predetermined variable on the instrument (Normal shift * Pre-leisure), controlling for normal shift time, pre-leisure day, and hospital, month, and year of birth fixed effects. Sample is restricted to single births, unscheduled C-sections and vaginal births that take place on working days.

and outcome probabilities are close to 0 or 1 (Chiburis et al., 2012; Bhattacharya et al., 2006; Nielsen et al., 2009). Given that we work in a low C-section rate setting and examine relatively rare outcomes, we expect bivariate probit to outperform 2SLS in terms of efficiency. In the results section, we report marginal effects for both estimators.[21]

[21] Bivariate probit models estimate unconditional average causal effects. In contrast, 2SLS estimates the LATE. However, in practice, the average causal effects produced by bivariate probit are likely to be similar to 2SLS estimates (Angrist and Pischke, 2009).

2.3.2.2 Differences-in-differences

Our second empirical strategy applies a differences-in-differences approach to a sample of sibling pairs. We restrict the sample to families where the older sibling was born by vaginal delivery and compare the health gap between siblings in families where the second child was born by an unplanned C-section against families where the second child was born by vaginal delivery. This enables us to control for all time-invariant unobserved heterogeneity at the family level and the effect of birth order. Our empirical strategy builds on numerous papers that have used siblings fixed-effects to estimate the impact of health shocks while in-utero or after birth (e.g. Oreopoulos et al., 2008; Almond et al., 2009; Almqvist et al., 2012; Aizer et al., 2016) and extends the model to a difference-in-differences specification with family fixed-effects. A related approach is used by Black et al. (2017) to study the impact of child disability on sibling outcomes.

We estimate the following equation:

$$Y_{if} = \psi_0 + \psi_1 Secondborn_{if} + \psi_2 Secondborn_{if} \times CS_{if} + X'_{if}\psi_3 + \gamma_f + \delta_m + \lambda_y + \phi_h + \eta_{if}, \quad (2.6)$$

where Y_{if} is the health outcome of child i in family f, $Secondborn_{if}$ is a dummy variable equal to 1 for the second child and 0 for the first child, CS_{if} is an indicator equal to 1 for unplanned C-section and 0 for vaginal delivery, X_{if} is a vector with the same pregnancy and maternal controls of equation (2.3), except for maternal characteristics that are time-invariant, and diagnoses during delivery (prolonged and obstructed labor)[22], γ_f, δ_m, λ_y and ϕ_h are family, month, year, and hospital of birth fixed effects, respectively.[23] We cluster standard errors at the family level. Our parameter of interest is ψ_2, which identifies

[22]We do not include these diagnoses during labor as controls in the IV specification, given that we find evidence that they can be an outcome of the time and type of day.

[23]We cannot estimate the baseline effects of the CS_{if} indicator, which are absorbed by the interaction $Secondborn_{if} \times CS_{if}$, since by construction of our sample only second children have C-sections.

the change in the health gap between siblings in families where the first child was born by vaginal delivery and the second child by C-section compared to families where both children were born by vaginal delivery.

We do not include families whose older child was born by C-section for two reasons. First, mothers who have a C-section in the first delivery and a vaginal birth in the second delivery are a very selected sample, given the very high probability of having a repeat C-section.[24] Second, some studies find that having a C-section is associated with lower fertility (Halla et al., 2016; Keag et al., 2018). We abstract from these concerns by focusing on mothers whose first birth was a vaginal delivery.

Even though our rich data sources make it possible to control for a large set of observable characteristics, it could be that there are sibling-specific unobservable differences that vary within family. In particular, younger siblings born by C-section could be negatively selected compared to their vaginally-delivered older siblings if the cesarean delivery is caused by complications, either during the pregnancy or delivery, that we cannot observe in our data. These unobservable complications could cause our estimates to be negatively biased. Thus, our difference-in-difference estimates could overestimate the impact of C-sections on the different diagnoses. In section 2.5.2 we assess the magnitude of the potential bias and provide evidence that it is relatively small. We will nonetheless keep the direction of this bias in mind when interpreting the results from this strategy.

[24] In 2010, The American College of Obstetricians and Gynaecologists (ACOG) encouraged doctors to allow women to opt for a vaginal delivery after a C-section, but the number of vaginal births after C-section has remained low (American College of Obstetricians and Gynecologists, 2010).

2.4 RESULTS

2.4.1 *Neonatal outcomes*

We first estimate the impact of C-sections on neonatal outcomes. Table 2.2 shows our OLS (first panel), 2SLS (second panel), bivariate probit marginal effects (third panel) and differences-in-differences (fourth panel) estimates. We find that the OLS results replicate existing findings. Cesarean sections are associated with adverse outcomes at birth and higher neonatal mortality.[25] Our 2SLS estimates are not significant for any of the outcomes. However, the magnitude of coefficients and large standard errors suggest that we cannot reject that there is a (potentially large) effect on neonatal outcomes. As discussed in section 2.3.2.1, 2SLS estimates are expected to be particularly uninformative with low treatment and outcome probabilities.

The bivariate probit coefficients are substantially more precisely estimated than the 2SLS results. Yet, all point estimates from the bivariate probit models are within the confidence intervals of the 2SLS estimates. The bivariate probit results suggest that unplanned C-sections increase the probability of having a low Apgar score (Apgar lower than 7), being admitted to the intensive care unit and receiving assisted ventilation. The magnitude of the bivariate probit estimates are similar to OLS estimates. However, we do not find significantly increased mortality risk within seven days after birth. The results from the differences-in-differences models give support to these findings with similarly-sized and more precise coefficients. Overall, our results suggest that unplanned C-sections have a negative impact on neonatal health. However, these adverse effects do not translate into a higher probability of early neonatal mortality.

[25]The OLS estimation is ran in a sample that only excludes planned C-sections and births for which we do not observe parity. The specification includes the full set of controls and fixed effects described in equation (2.1), as well as controls for birth order.

Table 2.2: Neonatal outcomes

	(1) Low Apgar 1	(2) ICU	(3) Assisted ventilation	(4) Neonatal mortality
OLS	0.068***	0.118***	0.027***	0.002***
	(0.001)	(0.001)	(0.001)	(0.000)
$\overline{Y}$	0.049	0.087	0.009	0.001
N	1119467	1120932	1120932	1119842
2SLS	-0.018	-0.088	-0.006	0.006
	(0.140)	(0.170)	(0.061)	(0.023)
$\overline{Y}$	0.066	0.106	0.012	0.002
N	392017	392560	392560	392173
Biprobit	0.104***	0.163***	0.017***	-0.001
	(0.008)	(0.009)	(0.005)	(0.005)
$\overline{Y}$	0.066	0.106	0.012	0.002
N	392017	392560	392560	392173
Diff-in-diff	0.053***	0.111***	0.036***	0.001
	(0.007)	(0.007)	(0.004)	(0.002)
$\overline{Y}$	0.038	0.070	0.006	0.001
N	644551	645292	645292	644746
First-stage F	24.996	25.216	25.216	26.007

This table shows the estimates of the effect of an unplanned CS on different neonatal health indicators by OLS, 2SLS, bivariate probit and differences-in-differences estimation (see equations (2.2), (2.3), and (2.6)). Specifications as detailed in section 2.3.2, with the full set of fixed effects and controls. Robust standard errors (in parentheses) in panels 1-3, and standard errors clustered at the family level in the differences in differences panel. First-stage F statistic from 2SLS and bivariate probit specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.4.2 *Later infant health*

We now turn to the results of the long-run effects of C-sections on health outcomes. Table 2.3 shows the OLS (first panel), two-stage least squares (second panel), bivariate probit (third panel) and differences-in-differences (fourth panel) marginal effect estimates at ages 5 and 10. We analyze health conditions that have been extensively documented in the literature as being positively associated with cesarean deliveries: type 1 diabetes, obesity, asthma, and other atopic diseases (atopic dermatitis and allergic rhinitis). Given that we study health outcomes for children who are born from 1990 to 2014, the sample size decreases as we consider older ages. We report year by year bivariate probit and diff-in-diff estimates up to age 15 in Figures 2.3 and 2.4, respectively. We report our OLS estimates in Figure 2.A.3.

The OLS estimates suggest that cesarean sections are associated with a higher probability of asthma, obesity and atopic diseases. These findings are consistent with existing studies that have documented significant associations between cesarean sections and metabolic and immune-related conditions. However, we do not detect that C-sections are associated with a higher probability of type 1 diabetes diagnosis.

The 2SLS results suggest that unplanned C-sections increase the probability of having a type 1 diabetes diagnosis before age 5, even though the effect is not significant by age 10. The effect size of the estimate is large, but very imprecise. Our results suggests 9 percentage point increase in the probability of type 1 diabetes, but are consistent with an increase ranging from 6.3 to 12.5 percentage points. The 2SLS estimates for asthma are not significant. However, the lack of precision does not enable us to rule out even very large (positive or negative) effects. For instance, the estimates by age 5 suggest that the impact of C-sections may range from -4.2 pp to 18.4 pp. Finally, the 2SLS estimates for obesity and atopic diseases are not significant, but also too imprecise to rule out very large effects.

Table 2.3: Child diagnoses by age

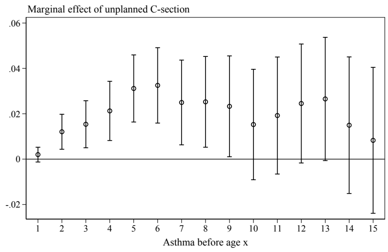
	Type 1 diabetes		Asthma		Obesity		Atopy	
By age:	5	10	5	10	5	10	5	10
OLS	0.000	0.000	0.007***	0.010***	0.000***	0.002***	0.002**	0.004***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
$\overline{Y}$	0.003	0.006	0.045	0.071	0.001	0.004	0.044	0.061
N	807035	556009	807035	556009	807035	556009	807035	556009
2SLS	0.089**	0.062	0.074	-0.121	0.001	0.000	0.022	0.110
	(0.036)	(0.044)	(0.113)	(0.139)	(0.013)	(0.034)	(0.112)	(0.127)
$\overline{Y}$	0.003	0.006	0.040	0.070	0.001	0.004	0.041	0.058
N	296998	217768	296998	217768	296998	217768	296998	217768
Biprobit	0.003	0.003	0.031***	0.015	0.001	0.003	-0.008	0.021
	(0.002)	(0.004)	(0.009)	(0.015)	(0.001)	(0.003)	(0.010)	(0.013)
$\overline{Y}$	0.003	0.006	0.040	0.070	0.001	0.004	0.041	0.058
N	296998	217768	296998	217768	296998	217768	296998	217768
Diff-in-diff.	-0.001	-0.001	0.014**	0.011	0.001	0.001	0.003	-0.001
	(0.001)	(0.003)	(0.006)	(0.009)	(0.001)	(0.002)	(0.005)	(0.007)
$\overline{Y}$	0.003	0.006	0.045	0.070	0.001	0.004	0.044	0.060
N	510075	366885	510075	366885	510075	366885	510075	366885
First-stage F	25.725	29.546	25.725	29.546	25.725	29.546	25.725	29.546

This table shows the estimates of the effect of an unplanned CS on the probability of the child having each diagnosis by age by OLS, 2SLS, bivariate probit and differences-in-differences estimation (see equations (2.2), (2.3), and (2.6)). Specifications as detailed in section 2.3.2, with the full set of fixed effects and controls. Robust standard errors (in parentheses) in panels 1-3, and standard errors clustered at the family level in the differences in differences panel. First-stage F statistic from 2SLS and bivariate probit specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

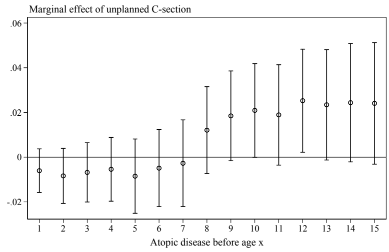
Similarly to our results for neonatal outcomes, the bivariate probit estimates (marginal effects) are substantially more precisely estimated than the 2SLS coefficients. Yet, practically all point estimates from the bivariate probit models are within the confidence intervals of the 2SLS estimates. For type 1 diabetes, the coefficient is much smaller than the coefficient from the linear model and not significant. For asthma, the results suggest a significant increase in the probability of a diagnosis by age 5 of 0.031 (95% CI 0.022–0.04). Even though estimates are noisier and no longer significant by age 10, the results in Figure 2.3 show that unplanned C-sections significantly increase the probability of an asthma

FIGURE 2.3: Bivariate probit estimation – Child diagnoses by age

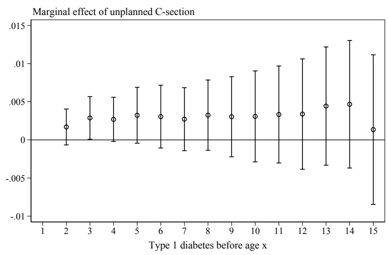
(a) Asthma



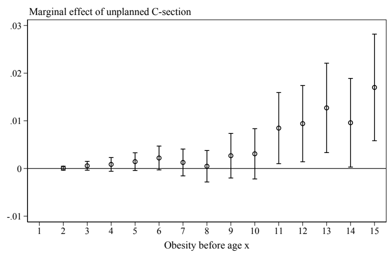
(b) Atopy



(c) Type 1 diabetes



(d) Obesity



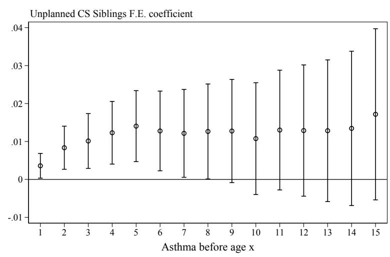
Notes: The figure plots the marginal effects from the bivariate probit estimation of the effect of unplanned CS on the probability of each diagnosis by age, with our usual specification. All regressions include hospital, year and month of birth fixed effects and the full set of controls as described in Section 2.3.2.1.

diagnosis for children as young as 2 years old. The effect is statistically significant up to age 9. For obesity, the bivariate probit results are precisely estimated at zero at age 5 (0.001, 95% CI 0.000–0.002) and 10 (0.003, 95% CI 0.000–0.006). However, the results in Figure 2.3 show a statistically detectable effect from age 11. Finally, we do not find a significant impact on atopic diseases at age 5 or 10.

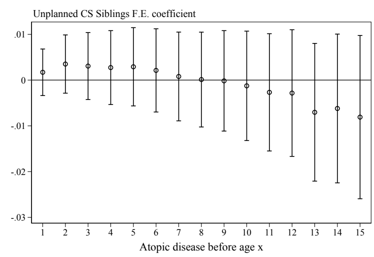
The differences-in-differences results are very similar to the bivariate probit results. We find that the second-born child has substantially greater risk of having an asthma diagnosis by age 5 than the first-born child in families where the second child is born by C-section. Similarly to the bivariate probit estimates, Figure 2.4 shows that this effect is significant from ages 1 to 8. Despite the fact that our differences-in-differences estimates could be negatively biased (Section 2.3.2.2), we do not find any significant effects on obesity, atopic diseases or type 1 diabetes. These results reinforce the conclusion that C-sections do not have impact on these outcomes.

FIGURE 2.4: Diff-in-diff analysis – Child diagnoses by age

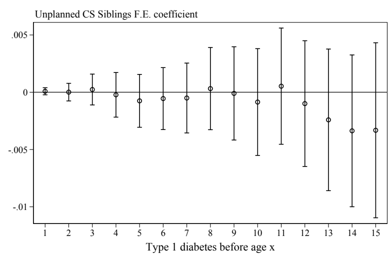
(a) Asthma



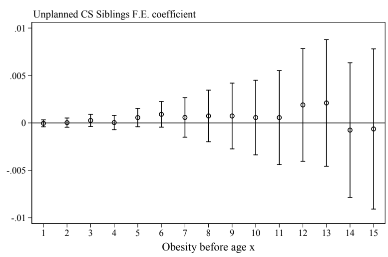
(b) Atopy



(c) Type 1 diabetes



(d) Obesity



Notes: The figure plots the coefficient of unplanned C-section for each diagnosis by age in family fixed effects models. All regressions include family, hospital, year and month of birth fixed effects and the full set of controls as described in Section 2.3.2.2.

Overall, our results suggest that unplanned C-sections increase the probability of suffering from asthma during childhood. The magnitude of this effect differs slightly depending on the estimation method. The bivariate probit estimates indicate a slightly larger but more imprecisely estimated impact (around 2 pp on average for ages 5 to 10) than the estimates based on differences-in-differences analysis (1.3 pp). By comparing these estimates to the sample mean, we find that the less precise bivariate probit estimates suggest a 36% increase in the probability of having asthma diagnosis (compared to the mean of 5.5% over ages 5-10), while the differences-in-differences estimates suggest a 21% increase (compared to the sample mean of 5.8%). The latter is closer to the 20% increase in the risk of asthma that is documented in recent meta-analyses (Thavagnanam et al., 2008; Keag et al., 2018).

Our analysis indicates that C-sections do not increase the probability of type 1 diabetes or atopic diseases. For diabetes, we can rule out effects larger than 0.7 pp at age 5 using the bivariate probit model and larger than 0.1 pp using the differences-in-differences model. For atopic diseases, in turn, our results discard effects larger than 1.2-1.3 pp with both methods. Finally, bivariate probit results suggest there might be an effect of C-sections on obesity after age 11. This observation is consistent with the evidence that puberty is a vulnerable period for the development of overweight and obesity (Lobstein et al., 2004). However, our analysis is not conclusive in this regard, as the results from the differences-in-differences estimation do not corroborate this finding. For younger ages, all methods suggest that there is no impact on obesity. For instance, estimates at age 5 enable us to rule out effects larger than 0.3 pp.

One potential limitation of our analysis is that we study diagnoses made at inpatient or outpatient visits to a hospital. For some outcomes, these diagnoses may be a good approximation to the true prevalence of the disease, while for others hospital diagnoses may lead to underestimation. A previous study on type 1 diabetes documents that in Finland practically all new type 1 diabetes diagnoses are made in a hospital and listed in the Hospital Discharge Register (Harjutsalo, 2008). This evidence implies that we are able to observe practically all type 1 diabetes diagnoses in our population of interest. However, since 1994, diagnoses for asthma in Finland are often made by general practitioners (Tuomisto et al., 2010). Thus, we are likely to trace only the most severe cases of asthma. The same might be true for atopic disesase and obesity.[26] In any case, OLS results show that C-sections in general are associated even with these hospital diagnoses. Our analysis thus highlights the importance of dealing with the endogeneity of the delivery mode.

2.5 VALIDITY CHECKS

2.5.1 *Exclusion restriction and sensitivity checks*

Our instrumental variables strategy relies on the assumption that the interaction of regular working hours and days that precede a weekend or public holiday affects health outcomes only through its impact on the likelihood of cesarean sections. We argue that, in this setting, this is likely to hold, since a violation would require other changes to happen on days that precede a public holiday but only during the regular shift. In the following, we provide several pieces of evidence that support the credibility of this assumption.

First, we explore the overall activity at maternity wards across the different types of days. Figure 2.A.4 (the first panel) shows the proportion of planned cesarean sections by time of birth and type of day. We find that scheduled activity is organized very similarly during all working days. Moreover, we compare the number of births by type of day and weekday (Figure 2.A.4, second panel) and do not find any evidence of maternity ward crowding during the days that precede a public holiday.

Second, we explore the quality of care provided during different weekdays. The first panel of Figure 2.A.5 shows that the probability of having a low Apgar score (below 7) does not differ between weekdays or type of day, suggesting that the quality of care during labor and

[26]There is some evidence that, among children, ICD-coding underestimates the true prevalence of obesity. ICD-coded cases have a higher BMI and higher healthcare utilization than those not coded (Kuhle et al., 2011).

Table 2.4: Validity checks

	Birth weight	Asthma at age 5 for sample: Thursdays vs Fridays	Asthma at age 5 for sample: Excluding inductions
Biprobit	-	0.023	0.036^{***}
	-	(0.015)	(0.010)
$\overline{Y}$	-	0.040	0.039
N	-	117826	246933
Diff-in-diff.	-5.416	-	0.017^{**}
	(7.617)	-	(0.007)
$\overline{Y}$	3566.117	-	0.044
N	645134	-	440291

This table shows, in column 1, a placebo regression where the outcome is birth weight; and in columns 2 and 3, the results from the bivariate probit (top) and the differences in differences (bottom) estimation of the impact of unplanned CS on the probability of asthma diagnosis by age 5 restricting the sample to births taking place on Thursdays or Fridays (col. 2) or to non-induced births (col. 3). Specifications as detailed in sections 2.3.2.1 and 2.3.2.2, with the full set of fixed effects and controls. Robust standard errors (in parentheses) for bivariate probit results, and standard errors clustered at the family level in the differences in differences panel.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

delivery does not differ by type of day. Figure 2.A.5 (second panel) shows the probability of early neonatal mortality, defined as death of a live-born baby within the first seven days of life, by weekday and type of day. We expect that this measure would capture changes in the quality of care. We do not find evidence that early neonatal mortality is higher for babies born on days that precede a public holiday compared to other weekdays. Moreover, we do not find that mothers who have a C-section on a day that precedes a public holiday have a longer length of stay than mothers who have a C-section on other weekdays (Figure 2.A.6). We interpret these findings as evidence that the quality of care remains constant across all working days.

Third, since babies born on days that precede a public holiday or weekend stay in the hospital during the following non-working days, one could argue that their quality of post-natal care is worse compared to children born on other working days. This would be constant for both babies born during the regular shift and at other times, and hence would not necessarily comprise the exclusion restriction. Yet, in what follows we assess this concern. Table 2.4 shows the coefficients for IV regressions that restrict the sample to babies born on Thursdays or Fridays.[27] We find, despite the reduced sample size, that the results from this estimation are consistent with our main results.

Finally, we report in Figure 2.2 that mothers who give birth during the regular working hours on days that precede a public holiday do not have higher probability of having induced labor. However, the induction of labor is likely to offer more room for discretionary behavior, in which case the decision to perform a C-section might be more sensitive to physician demand for leisure.[28] In other words, we expect that mothers whose labor has been artificially induced are more likely to be part of the complier population. Column 3 in Table 2.4 shows that our coefficients remain about the same if we exclude mothers whose labor was induced from our sample. The same conclusion holds if we exclude inductions from our differences-in-differences estimation. These results suggest that our findings are not driven by mothers whose labor has been induced after an admission to the maternity ward.

[27] The average length of stay in our sample is four days. The majority of babies born on Thursdays and Fridays are hospitalized during the weekend.

[28] Recent evidence casts doubt on the commonly-held belief that induction of labor increases the risk for cesarean delivery. In particular, recent studies show that inductions at full term do not increase the risk of cesarean delivery (Saccone and Berghella, 2015) or even lower it (Mishanina et al., 2014), with no increased risks for the mother and some benefits for the fetus.

2.5.2 *Differences-in-differences validity checks*

The results from our differences-in-differences model with family fixed effects could be biased if there are unobservable characteristics correlated with the mode of delivery that vary within family and across siblings. Under this scenario, this methodology would yield upward biased estimates. However, as shown in Section 2.4, our differences-in-differences results suggest that C-sections do not increase the risk of developing various immune-mediated diseases that have previously been associated with cesarean births.

To assess the extent to which these results could be explained by selection, we first run a regression using birth weight as a placebo outcome, given that it cannot be affected by unplanned C-sections. Table 2.4 shows that our differences-in-differences model with family fixed effects does not predict birth weight. This result supports the validity of this strategy: family fixed-effects, jointly with the large set of controls, seem to be taking into account general health differences between siblings born by C-section and vaginal delivery.

Second, we compare our differences-in-differences estimates to those from other samples of sibling pairs where we expect the second child to be negatively selected with respect to their older sibling, but where none of them was born by C-section. These samples include (i) a sample of siblings where the first child is born by eutocic birth and the second child is born either by eutocic or by instrumented birth, and (ii) a sample of siblings where the first born had a low-risk pregnancy and the second born had either a low- or a high-risk pregnancy, while all children in the sample were born by vaginal delivery.[29] Consequently, we assess the health gap between siblings across families that had a

[29] An eutocic delivery is a vaginal delivery with no instrumentation. We define a high-risk pregnancy as a pregnancy where the mother had at least one of these complications: a positive result in the glucose tolerance test, an hospitalization during pregnancy due to blood loss, hypertension, eclampsia or placenta previa. A low-risk pregnancy is defined as the absence of these issues.

complication during the second birth or during the second pregnancy compared to families where none of the siblings encountered any of these complications during pregnancy or birth.

Table 2.5: Validity of differences-in-differences

	Neonatal health				Diagnosis by age 5			
	Low Apgar	ICU	Assisted Ventilation	Neonatal mortality	Type 1 diabetes	Asthma	Obesity	Atopy
Instrumented	0.060***	0.020**	0.001	-0.001	-0.001	-0.004	-0.001	-0.006
	(0.009)	(0.009)	(0.003)	(0.002)	(0.002)	(0.008)	(0.001)	(0.009)
$\overline{Y}$	0.028	0.061	0.005	0.001	0.003	0.044	0.001	0.044
N	534119	534689	534689	534264	428392	428392	428392	428392
Risk pregnancy	0.001	0.016	0.003	0.001	-0.001	0.002	-0.000	0.005
	(0.007)	(0.010)	(0.004)	(0.002)	(0.002)	(0.009)	(0.001)	(0.009)
$\overline{Y}$	0.035	0.062	0.005	0.001	0.003	0.044	0.001	0.044
N	608688	609368	609368	608909	482536	482536	482536	482536

This table shows the results from sibling fixed effect models, following the specification in equation (2.6), for two different samples of children: in the top panel, for a sample of sibling pairs where the first child was born by eutocic birth, and the second child is born either by eutocic or instrumented vaginal birth; in the bottom panel, for vaginally delivered sibling pairs where the first child did not have a high-risk pregnancy and the second child had a low- or high-risk pregnancy. The top panel coefficient represents the change in the health gap between siblings in families where the second child was born by instrumented vaginal delivery, while the bottom panel coefficient represents the same for families where the second child had a high-risk pregnancy. All specifications include family, hospital, year and month of birth fixed effects and the controls described in section 2.3.2.2. Standard errors are clustered at the family level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5 shows our differences-in-differences estimates using these samples of siblings. The first four columns show that, compared to families where both siblings were born by eutocic birth, second children born by instrumented vaginal delivery have worse neonatal health than their older siblings who had an eutocic birth. We find a significantly higher probability of having low Apgar scores and of being admitted to the ICU (top panel). In the bottom panel, we can see that children who experienced a high-risk pregnancy do not have significantly worse neonatal health by any of the indicators, even though all coefficients have a positive sign. In the last four columns, we explore if negative selection leading to instrumented birth or risk pregnancy is associated with a higher probability of having any of the diagnoses we analyze in section 2.4. We do not find evidence that siblings born by instrumented vaginal delivery or those who had a high-risk pregnancy have an increased risk of type 1 diabetes, asthma, atopic diseases or obesity at age 5. These observations suggest that our differences-in-differences results for asthma are unlikely to be explained by negative selection.

2.6 CONCLUSIONS

This paper provides new evidence on the effects of avoidable cesarean sections on various short- and long-term health outcomes. We use a novel instrumental variable estimation strategy to overcome the potential endogenenity of birth mode and abstract from cases in which C-sections respond to a clear clinical indication. Our empirical strategy builds on the finding that unplanned C-sections are more common during regular working hours on Fridays and working days preceding public holidays. We complement this empirical strategy by estimating a differences-in-differences model with family fixed effects that compares the health gap between siblings in families where the second child was born by unplanned C-section with the health gap between siblings who were both born by vaginal delivery.

Our results suggest that C-sections have a substantial negative impact on neonatal health. However, these adverse effects are not severe enough to translate into a higher probability of increased neonatal mortality. Our long-run analysis follows children from birth to age 15 and investigates the impact of C-sections on four health outcomes that have been consistently associated with C-sections: type 1 diabetes, asthma, obesity, and atopic diseases. In contrast to the OLS estimates, our instrumental variable and differences-in-differences estimates show that unplanned C-sections do not have a significant effect on the probability of having a type 1 diabetes, obesity, or atopic disease diagnosis. However, we do find that being born by an unplanned C-section increases the probability of having asthma. This effect is detectable from ages 1-2 and of similar size to the associations reported by previous studies (Thavagnanam et al., 2008; Keag et al., 2018). Our results are consistent with the hypothesis that mode of delivery can affect the development of immune-related conditions, but suggest more nuanced effects of C-sections than previous work.

This paper provides first evidence on the long-term effects of unplanned C-sections that do not respond to a clear medical indication, using inpatient and outpatient data for all children born in Finland from 1990 to 2014. Although we are able to observe most of the cases of type 1 diabetes, for some diagnoses (asthma, atopic disease, and obesity) we might be only able to trace the most severe cases, given that these conditions are often treated by general practitioners. However, the fact that our OLS estimation, which includes a large set of controls, shows significant associations of cesarean birth with these outcomes, highlights the importance of dealing with omitted variable bias when analyzing the impact of mode of delivery. Future work should focus on analyzing the causal effect of C-sections on obesity and other metabolic disorders using primary care data and anthropometric measurements.

We make use of the detailed diagnosis data to show that variation by time and type of day can be a valid source of variation to investigate the impact of avoidable C-sections. First, we show that mothers who give birth at regular working hours on pre-leisure days are comparable in terms of a extensive list of pregnancy, health, and sociodemographic characteristics to mothers who give birth during these times on the rest of working days. Second, we show that during the normal shift on these pre-leisure days physicians make greater use of more discretionary diagnoses as justification for the C-section. We also show that these additional C-sections are not performed to mothers who are in the medical profession, and whose mode of delivery has been shown by the literature not to respond to doctors' incentives (Johnson and Rehavi, 2016).

All in all, our results suggest that the additional C-sections performed during regular working hours on pre-leisure days are not driven by medical factors. We provide this evidence in the context of Finland, one of the countries with the lowest C-section rate in the world (OECD, 2017). We would expect this variation to provide an even stronger source of identification in other countries with higher rates of medical interventionism during childbirth. Thus, this paper hopes to provide a solid base upon which future research on the effects of avoidable cesarean sections can be built.

2

APPENDIX

Appendix 2.A

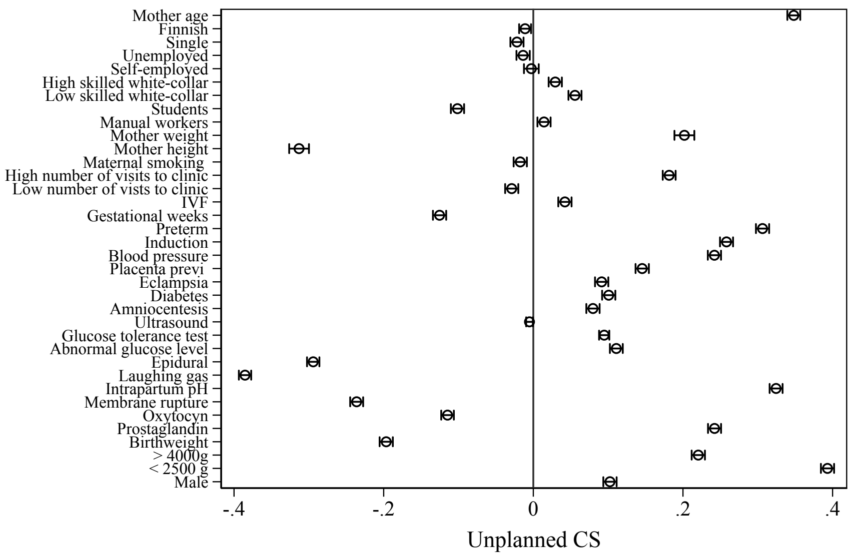


FIGURE 2.A.1: Difference in baseline characteristics by type of birth

Notes: The figure represents the coefficients and 95% CI from separate regressions of each (standardized) predetermined variable on an indicator taking value 1 if the mother had an unplanned C-section, and 0 if it was a vaginal delivery, controlling for normal shift time, pre-leisure day, and hospital, month, and year of birth fixed effects. Sample is restricted to single births, unscheduled C-sections and vaginal births that take place on working days.

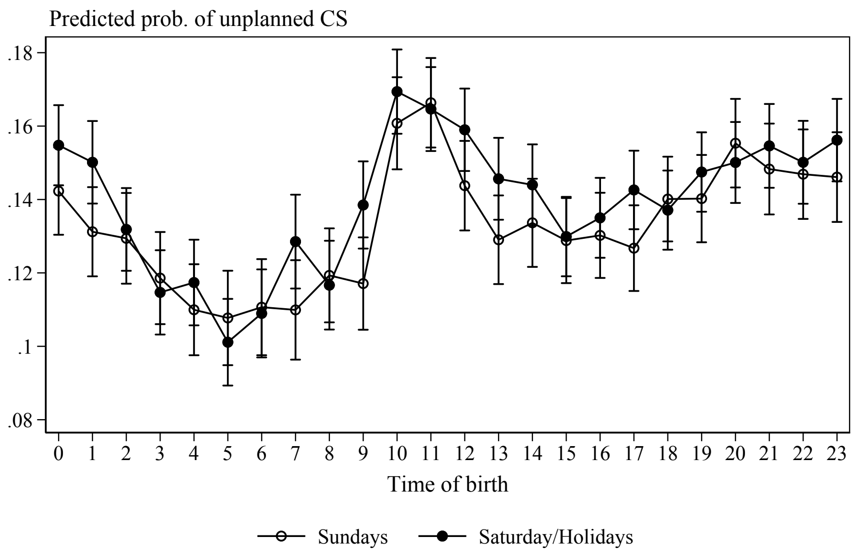
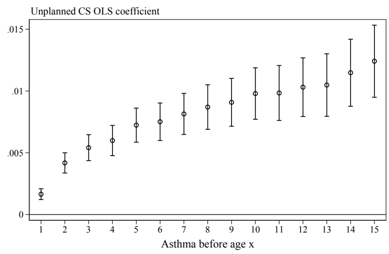


FIGURE 2.A.2: Predicted probability of unplanned C-section by time on weekends

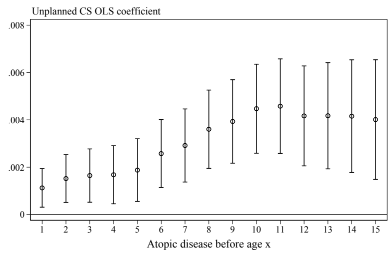
Notes: The figure represents the predicted probability of unplanned C-sections by time of birth for Sundays and for Saturdays or holidays, adjusting for hospital, month, and year of birth fixed effects. Sample is restricted to single births, unscheduled C-sections and vaginal births that take place on Saturdays or holidays and Sundays.

FIGURE 2.A.3: OLS estimation: Child diagnoses by age

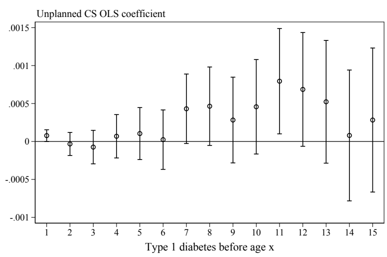
(a) Asthma



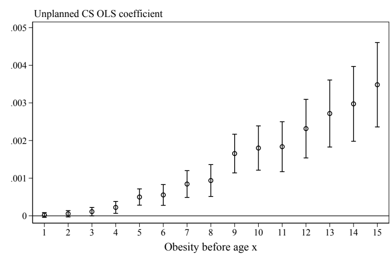
(b) Atopy



(c) Type 1 Diabetes



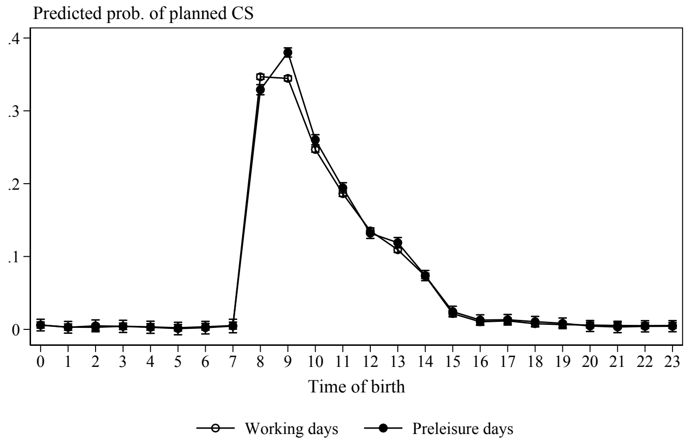
(d) Abesity



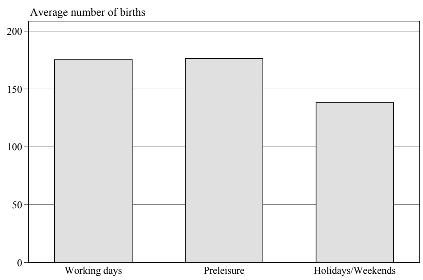
Notes: The figure plots the results from the OLS estimation of the effect of unplanned CS on the probability of each diagnosis by age, with our usual specification. All regressions include hospital, year, and month of birth fixed effects and the full set of controls described in section 2.3.2.

FIGURE 2.A.4: Activity at maternity wards by type of day

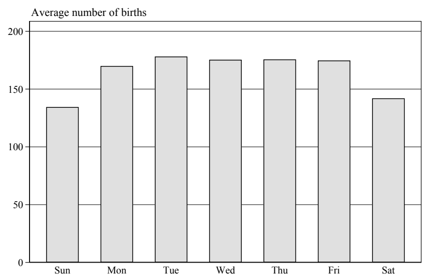
(a) Planned C-sections



(b) Births by type of day



(c) Births by weekday

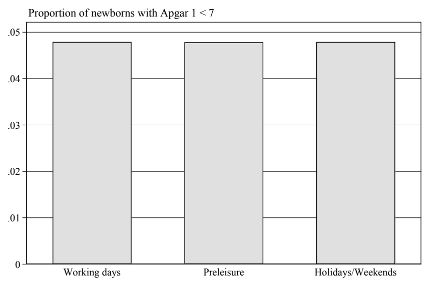


Notes: This figure plots, in the first panel, the probability of planned C-section by time of birth on pre-leisure working days and other working days, and in the second panel, the average number of births by type of day (column (a)) and by weekday (column (b)).

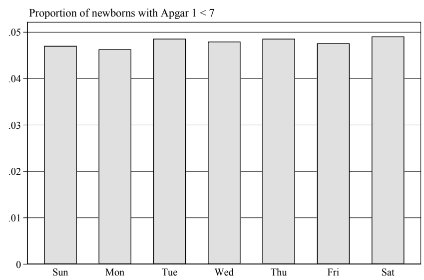
FIGURE 2.A.5: Quality of care by type of day

Low Apgar score

(a) By type of day

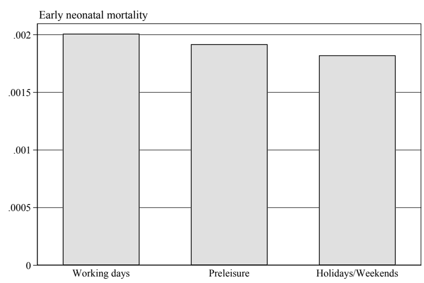


(b) By weekday

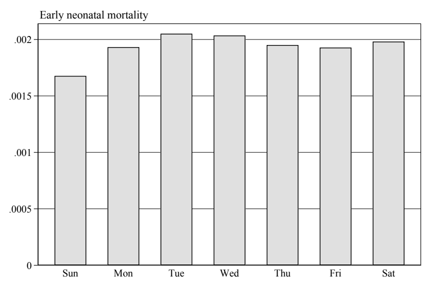


Early neonatal mortality

(c) By type of day



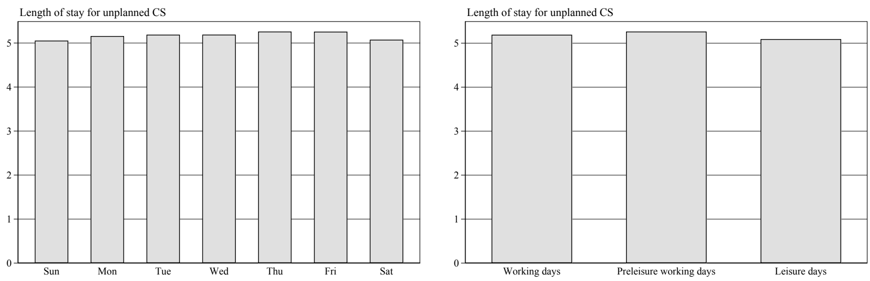
(d) By weekday



Notes: This figure plots, in the first panel, the probability of the newborn having low Apgar score and in the second panel the probability of early neonatal mortality by type of day (column (a)) and by weekday (column (b)). Sample is restricted to single births, unscheduled C-sections and vaginal births.

FIGURE 2.A.6: Mother length of stay by type of day

(a) By type of day (b) By weekday



Notes: This figure plots, in the left panel, the average length of stay of the mother for mothers who had a C-section by type of day, and in the right panel, by day of the week. Sample is restricted to single births and unscheduled C-sections.

Table 2.A.1: Summary statistics

	Full sample	
	Mean	SD
Background characteristics		
Mother's age	29.369	5.335
Finnish	0.958	0.200
Married	0.628	0.483
Unemployed	0.004	0.061
Selfemployed	0.017	0.128
High skilled white-collar	0.178	0.382
Low skilled white-collar	0.433	0.496
Student	0.095	0.294
Manual workers	0.180	0.384
Pregnancy characteristics		
Mother weight	66.780	14.033
Mother height	165.562	6.032
Tobacco during pregnancy	0.128	0.334
High visits clinic	0.239	0.426
Low visits clinic	0.190	0.392
IVF	0.003	0.057
Gestational weeks	39.702	1.853
Preterm	0.056	0.230
Previous CS	0.099	0.299
First birth	0.410	0.492
Blood pressure hospitalization	0.033	0.178
Placenta previa	0.003	0.052
Eclampsia	0.000	0.022
Gestational diabetes	0.007	0.085
Amniocentesis	0.029	0.168
Ultrasound	0.458	0.498
Glucose Tolerance Test	0.183	0.387
Glucose Tolerance Test Positive	0.049	
Childbirth characteristics		
Induction	0.165	0.372
Epidural	0.326	0.469
Laughing gas	0.453	0.498
Intrapartum pH	0.042	0.201
Membrane rupture	0.448	0.497
Oxytocyn	0.401	0.490
Prostaglandin	0.076	0.265
Birth weight	3520.736	571.55
Male	0.511	0.500
Mode of delivery		
Planned CS	0.071	0.257
Unplanned CS	0.101	0.301
Eutocic	0.763	0.425
Ventose	0.066	0.248
Forceps	0.001	0.033
Breech vaginal	0.005	0.073
Observations	1482884	

Table 2.A.2: Long-term outcome variables

Outcome	ICD-10 codes	Description
Asthma	J45, J46	Asthma is the most common chronic disease in children (Asher and Pearce, 2014). Asthma is an inflammatory disorder characterized by recurrent attacks of breathlessness and wheezing and can also cause cough, particularly in children. Recurrent asthma symptoms frequently cause sleeplessness, daytime fatigue, reduced activity levels and school and work absenteeism.[a] It is caused by a complex combination of genetic and environmental factors.
Atopic diseases	L20, J30.1-30.4, J30.8, J30.9	It includes atopic dermatitis and allergic rhinitis. Atopy is a predisposition toward developing certain allergic hypersensitivity reactions. Atopic dermatitis is a chronic inflammatory skin disease associated with cutaneous hyperreactivity to environmental trigger. It is believed to be the product of interactions between susceptibility genes, the environment and immunologic responses (Leung et al., 2004). Allergic rhinitis is characterized by one or more symptoms including sneezing, itching, nasal congestion, and rhinorrhea (Skoner, 2001).
Type 1 Diabetes	E10	Type 1 diabetes is a chronic auto-immune mediated disease. The body destroys beta cells, which are cells located in the pancreas that produce and segregate insulin, the hormone that regulates glucose levels in the blood. In type 1 diabetes patients, the body is unable to regulate glucose levels. This disease develops in genetically susceptible individuals, but the medical literature has recognized environmental factors as crucial in the triggering and development of the condition (Knip and Simell, 2012).
Obesity	E65-E68	It includes obesity, overweight, localized adiposity and other hyperalimentation. Obesity is defined as abnormal or excessive fat accumulation that may impair health. The prevalence of overweight and obesity among children and adolescents aged 5-19 has risen dramatically from just 4% in 1975 to just over 18% in 2016.[b] Although obesity is most commonly caused by excess energy consumption (dietary intake) relative to energy expenditure, the etiology of obesity is highly complex and includes genetic, physiologic, environmental, psychological, social and economic factors (Wright and Aronne, 2012). Recent research highlights the role of gut microbiota in the development of obesity (Ottosson et al., 2018).

[a] http://www.who.int/respiratory/asthma/en/
[b] http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Table 2.A.3: Public Holidays in Finland (Year 1992)

Public holiday	Date (1992)	Weekday (1992)
New Year's Day	January, 1	Wednesday
Epiphany[a]	January, 6	Monday
Good Friday[b]	April, 17	Friday
Easter Sunday[c]	April, 19	Sunday
Easter Monday[d]	April, 20	Monday
May Day	May, 1	Friday
Ascension Day[e]	May, 28	Thursday
Whit Sunday[f]	June, 7	Sunday
Midsummer Eve[g*]	June, 19	Friday
Midsummer Day	June, 20	Saturday
Finnish Independence Day	December, 6	Sunday
Christmas Eve[*]	December, 24	Friday
Christmas Day	December, 25	Saturday
Boxing Day	December, 26	Sunday

[a] Epiphany was moved to January 6 in 1992. Previously, Epiphany was the Saturday following January 5. [b] Moveable Friday before Easter Sunday. [c] Moveable Sunday following the first full moon on or after March 21. [d] Moveable Monday after Easter Sunday. [e] Moveable Thursday 39 days after Easter Sunday. Until 1992, the Ascension Day was the Saturday before the Thursday1. [f] Moveable Sunday 49 days after Easter Sunday. [g] First Friday on or after June 19. [*] No legal status as a public holiday, but included in collective labor agreements.

Table 2.A.4: Relation of the instrument with discretionary diagnoses vs. medical emergencies

	(1) Dystocia	(2) Suspected fetal suffering
Preleisure day	-0.002**	-0.001
	(0.001)	(0.001)
Normal shift	-0.002**	0.005***
	(0.001)	(0.001)
Normal shift*Preleisure	0.005**	0.002
	(0.002)	(0.002)
Observations	392560	392560
Adjusted R^2	0.074	0.057
Controls	YES	YES
F-statistic	9.211	0.607

This table shows the results from our usual first-stage specification, but with the following dependent variables: in column 1, an indicator for prolonged or obstructed labor; in column 2, an indicator equal to 1 if fetal scalp pH measurements were taken during labor. All specifications include hospital, year, and month of birth fixed effects, and the full set of controls as described in equation (2.2). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.A.5: First stage – Medical Professional Mothers vs. Others

Sample:	All non-medical mothers		Non-medical mothers with university education		Medical mothers	
Unplanned CS	(1)	(2)	(3)	(4)	(5)	(6)
Normal shift	0.014***	0.017***	0.014***	0.018***	0.025***	0.025***
	(0.001)	(0.001)	(0.002)	(0.005)	(0.006)	(0.005)
Preleisure day	0.001	-0.002	-0.001	-0.002	0.000	-0.002
	(0.002)	(0.002)	(0.003)	(0.007)	(0.007)	(0.007)
Normal shift*Preleisure	0.016***	0.015***	0.015***	0.014***	-0.004	-0.003
	(0.003)	(0.003)	(0.005)	(0.012)	(0.012)	(0.012)
Observations	367825	367825	147463	147463	22526	22526
Adjusted R^2	0.008	0.071	0.008	0.072	0.006	0.068
Controls	NO	YES	NO	YES	NO	YES
Mean of Y	0.146	0.146	0.152	0.151	0.154	0.154
First-stage F	28.998	27.378	10.428	9.609	0.092	0.067

This table shows the usual first stage, with unplanned C-section as dependent variable, for different groups of mothers: all mothers not in the medical profession (columns 1-2), for mothers not in the medical profession with university education (columns 3-4), and for mothers in the medical profession (5-6). Medical mothers include doctors, nurses and midwives. All specifications include hospital, year, and month of birth fixed effects and the full set of controls as described in equation (2.2). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3

THE CAREER COSTS OF CHILDREN'S HEALTH SHOCKS

3.1 INTRODUCTION

Economists have long been interested in understanding the relationship between income and health (Deaton, 2013). The detrimental effect of health shocks on an individual's own labor market outcomes is well documented.[1] However, we know much less about the potential spillover effects that health shocks have on other family members.

In particular, the illness of a child is a stressful event that can have major implications for the well-being of the whole household. Families can incur substantial costs when deciding how to best cope with these health shocks and their associated long-term burden. Parents may need to increase the time spent caring for their child and decrease their labor supply. However, they may also face direct treatment costs, resulting instead in an increase in their labor supply.

[1] This includes, among others, papers by Bound et al. (1999); Cai et al. (2014); Dobkin et al. (2018); García-Gómez (2011); García-Gómez et al. (2013); Jones et al. (2019); Lindeboom et al. (2016); Lenhart (2019); Meyer and Mok (2019); Trevisan and Zantomio (2016); Wagstaff (2007)

Suffering a hospitalization during childhood is a situation dealt with by a relatively large number of families. For example, the rate of hospital admissions in 2009 in the United States among children aged one to seventeen was 30 per 1,000 children (Wier et al., 2011). Looking at the cumulative incidence of hospitalization, 36% of children born in Finland in 1990 suffered a hospitalization between ages one and eighteen, while 9.9% needed a long inpatient stay (over four days). Although child mortality rates are very low across European countries,[2] the death of a child can lead to significant emotional distress and can have enormous impacts on parents' well-being and labor market outcomes. However, our knowledge on how children's health shocks (both non-fatal and fatal) impact the economic well-being of families is surprisingly limited.

This paper contributes to bridging this gap by providing new evidence on the causal impact of a child's health shock on parental outcomes. I examine the effects of both severe hospitalizations and fatal health shocks on parents by leveraging long panels of high-quality administrative data from Finland on families' health and labor market trajectories. I exploit variation in the timing of health shocks among families whose child had a first severe health shock after school-starting age. Identification comes from comparisons of parents and children in the same respective age cohorts, but whose children experienced the health shock at different ages. I show that these families have very similar characteristics and were following very similar trends before the shock.

With this data and design, I provide precise causal estimates of parents' labor supply responses to children's health shocks or mortality. Using an event study approach, I first show that there is no indication that parents' outcomes deviate from the trend prior to the health shock

[2]On average, fewer than 1.25 deaths per 1,000 children between ages 5 and 14. United Nations data for 2018, available at: https://childmortality.org/data. Accessed September 2019.

of the child. For all outcomes, sharp breaks in the trajectories become visible just after the event. Overall, I find that the earnings of both mothers and fathers suffer a substantial and persistent drop after a serious health shock or the death of their child. Five years after a severe hospitalization, maternal earnings are 7.5% lower compared to the period prior to the shock. For fathers, earnings drop by 2.5%. At the extensive margin, these shocks also impact a mother's probability of being employed. In contrast, I do not find evidence of any effect on the probability of a father continuing to work. I also show that the effect is driven by health shocks that require persistent care, as measured by the number of hospital visits in the year after the shock. Furthermore, I find that three years after the death of a child, a mother's earnings have dropped by 23%. For fathers, the estimated coefficients are negative and large in magnitude, but imprecise, suggesting that fatal shocks may also have an impact on their earnings.

I exploit the richness of the data to explore several potential mechanisms. I do not find evidence that mothers switch jobs to more family-friendly firms after the shock. Nor do I find changes in the risk of marital dissolution. However, I do find that children's health shocks affect the mental well-being of parents, as measured by the number of visits to a specialist or hospital admissions with a primary mental health diagnosis. My results suggest that the impact of a child's severe hospitalization on parents' earnings might result from the combination of the increased time needed to care for the child and worsening of parents' mental health.

This paper contributes to several strands of the literature, including that on the effects of adverse health shocks on labor market outcomes. Numerous studies have analyzed the relationship between health shocks and income, though most focus on the impact of health shocks on the individual's own labor market outcomes (e.g, Bound et al., 1999; Cai et al., 2014; Dobkin et al., 2018; García-Gómez, 2011; García-Gómez et al., 2013; Jones et al., 2019; Lindeboom et al., 2016;

Lenhart, 2019; Meyer and Mok, 2019; Trevisan and Zantomio, 2016; Wagstaff, 2007). Using an event study approach, Dobkin et al. (2018) examine the economic consequences of hospitalizations for adults in the US. They find that earnings drop by 20% three years after a hospitalization. Meyer and Mok (2019) use survey data from the US and estimate a similar drop in earnings ten years after the onset of a disability.

Other studies have examined the spillover effects of health shocks, with particular attention paid to how one spouse's health shock affects the other spouse's employment and earnings.[3] Fadlon and Nielsen (2017) analyze the impact of a spouse experiencing a fatal or severe non-fatal shock on household labor supply. Using administrative data from Denmark and exploiting event studies together with a dynamic difference-in-differences approach, they find that fatal health shocks lead to an increase in the labor supply of the surviving spouse. In contrast, they do not find any significant response following a non-fatal health shock.[4] García-Gómez et al. (2013) explore the spillover effects of an acute hospitalization using data from the Netherlands. They find gender asymmetries in the response to a spouse's health shock: while wives are more likely to continue—or even start—working when their husbands fall ill, husbands are more likely to withdraw from the labor force when their wives fall ill. Jeon and Pohl (2017) use administrative data from Canada and observe a significant decline in the employment and earnings of individuals whose spouses are diagnosed with cancer.

Rellstab et al. (2019) instead examine the spillover effects of an older parent's unexpected hospitalization[5] on their children's labor supply. Utilizing a difference-in-differences model and administrative data from the Netherlands, they do not find significant effects on either employment or earnings. Black et al. (2017) exploit a difference-in-

[3]See, for example, García-Gómez et al. (2013); Fadlon and Nielsen (2017); Jeon and Pohl (2017); Jiménez-Martín et al. (1999)

[4]They use heart attacks and strokes as severe non-fatal health shocks.

[5]They exploit diagnoses classified by physical expert opinion as being unexpected hospitalisations, and thus plausibly exogenous.

differences approach and show that having a sibling with a disability has a negative spillover effect on children's test scores.

I build on this literature by providing the first causal evidence of the spillover effects of a child's health shock on the labor supply and mental well-being of the family. Several previous studies find a negative association between childhood disability or illness and maternal employment (e.g, Wasi et al., 2012; Wolfe and Hill, 1995).[6] A few papers make use of panel data and try to control for previous employment situation (Baydar et al., 2007; Burton et al., 2017; Kvist et al., 2013; Powers, 2003; van den Berg et al., 2017)[7]. However, children's health status is unlikely to be randomly distributed across families, meaning that families whose children have worse health are likely to be different from other families. This makes it difficult to distinguish between the effect of having a child with an illness and that of other confounding characteristics on maternal employment.

This paper advances existing knowledge by using high-quality administrative data combined with a research design that allows me to exploit precisely and objectively identified- health shocks, as well as to focus on a sample of families that are similar, differing only in the age at which their child suffered the health shock. Moreover, I show that the effect is only visible if the health shock imposes a substantial and persistent care burden on the parents.

My study also speaks to the literature that investigates the impact of parenthood on family labor supply, which shows sizeable effects on mothers' labor supply and earnings.[8] The most recent studies estimate

[6]Stabile and Allin (2012) review previous research and conclude that, taken together, the studies suggest that having a child with disabilities is associated with a higher likelihood that the mother (and less often the father) will either reduce their working hours or stop working altogether.

[7]van den Berg et al. (2017) make use of longitudinal data and match parents whose child died in a non-intentional accident to parents who did not lose a child. They find that mothers' annual earnings decrease by 12.5% on average, while fathers' earnings decrease by 8.8%.

[8]This includes, among others, papers by Adda et al. (2017); Angrist and Evans

that women's earnings decrease considerably following the birth of the first child, an effect that is persistent. The so-called child penalty[9] amounts to around 20% over the long run in Nordic countries (Kleven et al., 2019b; Sieppi and Pehkonen, 2019), between 30% and 45% in the United Kingdom and the United States, and as high as 50%-60% in Germany and Austria (Kleven et al., 2019a). Snaebjorn and Steingrimsdottir (2019) find that the child penalty is larger in families in which a child is born with a disability: affected mothers earn 13% less in the long run, while affected fathers earn 3% less.

I show here that beyond the costs of having a child, health shocks during middle childhood to teenage years also have a substantial impact on parents' labor market outcomes. In line with studies on the impact of children, my results suggest that the negative impact of children's health shocks is greater on women's earnings than on men's.

The paper is structured as follows: Section 3.2 lays out the empirical strategy and Section 3.3 provides background information about the institutional context and introduces the data. Section 3.4 reports the main results. Section 3.6 presents additional evidence to support the main conclusions. Section 3.7 explores the mechanisms of the effects. The final section concludes.

3.2 EMPIRICAL STRATEGY

I aim to analyze the impact of a child's health shock on parents' labor market outcomes and well-being. Child hospitalizations are unlikely to be randomly distributed, meaning that the characteristics and trajectories of families whose child suffers a health shock may be different from

(1998); Angelov et al. (2016); Benard et al. (2007); Bertrand et al. (2010); Bronars and Grogger (1994); Bütikofer et al. (2018); Fernández-Kranz et al. (2013); Hotz et al. (2005); Lundberg and Rose (2000); Lundborg et al. (2017); Paull (2008); Miller (2011); Sigle-Rushton and Waldfogel (2007); Waldfogel (1998)

[9] Defined as the percentage by which women fall behind relative to men due to having children.

other families. Figure 3.A.1 plots the coefficients of regressing different family and child characteristics on a dummy equal to 1 if the child suffered a severe hospitalization. Having a child who was hospitalized predicts almost all characteristics, suggesting that these families are different from others. Therefore, comparisons between these groups of families are likely to yield biased estimates of the causal impact of children's health shocks.

In order to overcome the potential endogeneity of children's health shocks, I exploit variation in their timing. Focusing on parents who have been exposed to a child's health shock at some point, I exploit variation in the age at which the child experienced the shock, conditional on the age of the parents and children. I focus on families whose child experienced a first shock after school-starting age to ensure that the mother's earnings follow parallel trends.[10]

I provide visually clear results of my estimation by utilizing an event study approach with a specification that follows recent work by Kleven et al. (2019b) and Nix and Andresen (2019), among others. In particular, I estimate the coefficients of indicator variables for years relative to the event ("event time"). For each parent in the dataset, the year of the shock is normalized to t=0 and all years are indexed relative to it. I construct a balanced panel of parents with observations dating from five years before and after the health shock. I run the following regressions for mothers and fathers:

[10]Given that mothers experience a sharp drop in their earnings and the probability that they work after childbirth, their earnings trajectories follow different trends during the first years after having had a child. Sieppi and Pehkonen (2019) replicate the analysis of Kleven et al. (2019b) for Finland and find that the child penalty stabilizes from age 6 onwards, which is also the school-starting age in Finland (Finnish National Agency for Education, 2018). In Figure 3.A.4, I show that families whose child suffers a severe hospitalization after school-starting age have very similar earnings trajectories prior to the health event.

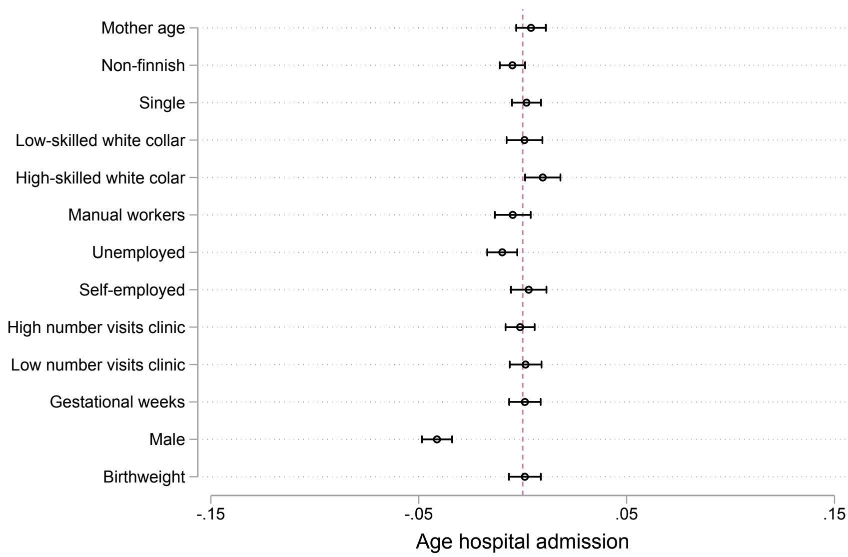
$$Y_{is} = \alpha + \sum_{t\neq -1, t=-5}^{t=5} \gamma_t \times I_t + \lambda A_{is} + \phi CBY_i + \omega_s + \varepsilon_{is} \tag{3.1}$$

where Y_{is} is the outcome of interest for individual i in calendar year s, which is regressed on event time dummies (I_t), parental age dummies (A_{is}), child's year of birth dummies (CBY_i), and calendar year dummies (ω_s). The event time dummy at $t = -1$ is the omitted category, meaning that the event time coefficients measure the impact of a child's health shock relative to the year prior to the event. Event time t runs from -5 to +5 years. The inclusion of the full set of age dummies controls non-parametrically for underlying life-cycle trends in parental outcomes, the calendar year dummies take into account time trends, while the child's birth year dummies control for cohort-specific effects.

Note that the variation in exposure to the health shock, or event time, arises from the age at which the child suffers the health shock, conditional on their year of birth, calendar year, and parents' age. Therefore, if the exact timing of the health shock is uncorrelated with the counterfactual outcome, conditional on parents' age profiles, the child's year of birth, and calendar-year fixed effects, the estimates can be given a causal interpretation as the impact of a child's health shock on earnings. Examples of scenarios that would violate this assumption are hospital admissions caused by a worsening of parents' earnings or simultaneous shocks that impact both parents' earnings and children's health.

One might expect the timing of the child's hospitalization to be related to their parents' earnings trajectories. I provide evidence to support that the change in maternal earnings is driven by children's health shocks. First, I demonstrate that comparisons between affected families eliminate most differences in observable pre-health shock characteristics, in contrast to comparisons between affected and unaffected families. Figure 3.1 plots the coefficients of regressing different maternal and child characteristics on children's age at hospital admission, controlling for child's year of birth fixed effects. The results indicate that families whose children experience a severe hospitalization at different ages have very similar observable characteristics prior to the health shock.[11] This is true for both non-fatal and fatal shocks.[12]

FIGURE 3.1: Differences in characteristics: within affected families



Notes: The figure shows the coefficients and 95% CI from separate regressions of each (standardized) variable with respect to children's age at hospital admission. All specifications include year of birth fixed effects. Standard errors are clustered at the mother level.

Second, in Figure 3.A.4, I show that families whose child suffered a severe hospitalization at different ages have very similar earnings

[11] Although boys and girls differ in their average age at hospital admission, columns (3) and (4) in Table 3.A.2 show that my results are robust to controlling for the child's gender.

[12] Results for fatal shocks can be found in Figure 3.A.2 and Figure 3.A.3.

trajectories prior to the health event, and that earnings did not decrease during the year prior to the event.[13] Moreover, the main advantage of using a non-parametric event study is that it allows a visual inspection of whether there was a trend in the outcome variable prior to the event. I do not find any evidence of an anticipatory drop in earnings (or in any other outcomes) before the event. Third, in Section 3.5, I show that the effect of a health shock is only visible if the child requires substantial and persistent care subsequent to hospitalization, as measured by the number of specialist visits and later hospital admissions. Finally, in Section 3.7, I explore two plausibly exogenous health shocks that have very different implications in terms of the care burden imposed on parents. I show that parental earnings do not respond to a severe health shock that, in general, does not require additional treatment (appendicitis), while there is a substantial drop following a severe hospitalization due to a more serious condition (cancer).

3.3 INSTITUTIONAL SETTING AND DATA

3.3.1 *Institutional Setting*

Finland has universal public health coverage. While primary health care is provided by local authorities in health centers, specialized medical care, consisting of specialist examinations and treatment, usually requires a physician's referral and is mainly provided in hospitals.[14] Emergency medical services, which involve treating acute illnesses or injuries, are also provided by hospitals.

There are twenty hospital districts in Finland, each of which has a central hospital. Hospital districts must provide a 24-hour emergency

[13]Figure 3.A.4 plots the different pre-health shock earnings trajectories for families whose child suffered an adverse health event at various ages. Families whose child suffered the shock after school-starting age follow very similar trajectories. However, this is not the case if the child experienced the shock at younger ages.

[14]The most common specialized medical care services are also available at some health centers.

medical service for dealing with urgent cases. Hospitals also offer specialized medical care on a 24-hour emergency basis. In many municipalities, hospitals also cover the emergency duties of health centers at night and during weekends (Ministry of Social Affairs and Health, 2013).

The private healthcare sector in Finland is relatively small, but has gained importance in recent years. There are only a few such hospitals, but private provision of specialist outpatient care is much more common (OECD, 2017). Although the use of private services is mainly financed through out-of-pocket payments, patients are eligible for a reimbursement from the National Health Insurance scheme. However, the effective reimbursement rate is only around 30% of the costs (Tynkkynen et al., 2016). In 2013, around one fifth of the Finnish population was covered by voluntary private health insurance, with children making up almost half of insured individuals (Tynkkynen et al., 2016).

In Finland, parents of ill children are entitled to different types of financial aid. First, during hospital treatment and subsequent care at home, parents can be granted the Special Care Allowance.[15] This aid is intended to compensate for lost income during the time that the child is undergoing medical treatment. Second, for disabled or chronically ill children, parents can be granted a disability allowance. The entitlement and the amount of the allowance are determined on the basis of the care, attention, and rehabilitation that the child requires. The payment period also depends on the assessment of how long care will be needed due to the illness or disability. Finally, family members can also be granted an informal care allowance if they take care of a severely disabled or chronically ill child at home.[16]

Families who face the death of a child are not entitled to receive any

[15]For a parent to be granted the Special Care Allowance, the attending physician must issue an statement confirming the seriousness of the illness and the need for the parent to participate in the child's care and treatment.

[16]Information available at: https://www.kela.fi/web/en/if-a-child-gets-ill.

allowance. Survivors' pension only replaces lost income when a family wage earner dies.

3.3.2 *Data*

I use rich individual-level administrative data from several sources to link family members. In particular, I merge employer-employee data from the Finnish Longitudinal Survey (FLEED-FOLK) with birth register data to identify families. I focus on the first child in each family.

The Finnish Medical Birth Register includes data on all live births from 1987 to 2014. The FLEED-FOLK records provide information for the entire population (aged between 16 and 70) from 1988 to 2015, with information on year of birth, education level, annual labor earnings, and employment status.

For health data, I use two different sources. The first is the Finnish Hospital Discharge Register, which contains information on diagnosed medical conditions and the exact date of diagnoses. This register contains all inpatient consultations in Finland from 1988 to 2015. From 1998 onwards, it also includes all outpatient visits to hospitals. All diagnoses are recoded using the International Classification of Diseases (ICD) system.[17] The second dataset is the Cause of Death Registry, which includes information on all death dates and causes between 1990 and 2015.[18]

I analyze two different health shocks: severe hospitalizations and fatal health shocks. Severe hospitalizations are defined as admissions resulting in a stay that lasts longer than 75% of all hospital stays,[19] which is equivalent to a four-day stay or longer.[20] For severe hospital-

[17]Diagnoses from between 1987 and 1995 are recorded using the ICD-9 classification. Those from 1996 onwards are recorded using ICD-10 classification.

[18]The statistics on causes of death are compiled based on the 10th revision of the International Classification of Diseases (ICD-10)

[19]Figure 3.A.5 shows the distribution of the length of hospital stays.

[20]In Section 3.6.2, I show that my results are robust to using different definitions for severe hospitalizations.

izations, the sample includes families whose first-born child suffered a first inpatient stay in an acute care hospital between ages six and eighteen. For fatal shocks, the sample consists of all families whose first-born child died between ages six and eighteen.[21]

Subsequently, I analyze parents' earnings and employment status five years before and after their child's health shock. In particular, my outcome variables are annual earnings in euros and the probability of being employed during each year. I also explore post-transfer earnings, which include any transfers from public or private sources. In order to explore the mechanisms behind changes in earnings, I further investigate the number of visits to a hospital or mental health specialist, the probability of divorce, the probability of working in a state-owned enterprise, and the probability of the mother changing employers in a given year.

I do not impose any restrictions in terms of parents' relationship status. They may be separated, divorced, or not in a relationship. In the latter case, I do not observe the father, meaning that my sample contains more mothers than fathers.

Table 3.A.1 shows summary statistics for the final samples used in the analysis. In Finland, 25,960 children were admitted to the hospital between ages six and eighteen during the period 1996 to 2008. Of these, 8,546 were severe hospitalizations. During these years, there were 358 child deaths in Finland. In Figure 3.A.7, I show some additional descriptive statistics for a specific cohort that can be followed until adulthood (children born in 1990); almost 6% of these children suffered a severe hospitalization between ages 6 and 18, while 0.2% suffered a fatal shock.

[21] Figure 3.A.6 shows the number of observations for each age between six and eighteen years. Severe hospitalizations and fatalities show considerable variation in terms of the age at which they occur.

3.4 RESULTS

3.4.1 *Severe Hospitalizations*

Figure 3.2 presents the event study estimates of the impact of a child's severe hospitalization on maternal labor market outcomes (Table 3.A.2 provides further detail about the estimates).[22] There is no indication that maternal earnings or probability of employment follow a different trend before the child's hospital admission. For all outcomes, a sharp break in the trajectories becomes visible just after the event. This lends support to the identifying assumption that the timing of the child's hospital admission is uncorrelated with the counterfactual outcomes.

Panel (a) in Figure 3.2 shows the results for maternal labor earnings in euros.[23] A child's severe hospitalization causes a significant and persistent drop in maternal earnings. Five years after the shock, mothers earn about 1,200 less than one year before the event. Compared to mean earnings the year before the event, the magnitude of the effect is substantial: the loss in income amounts to a 7.5% decrease in maternal earnings. Panel (b) plots the results for probability of employment. There is a drop in the probability of working, which is only significant at the time of the shock and one and three years after. Three years after the shock, the probability of working is 1.8 percentage points lower. This amounts to a 2.5% decrease in a mother's working probability with respect to the mean level of employment prior to the event.

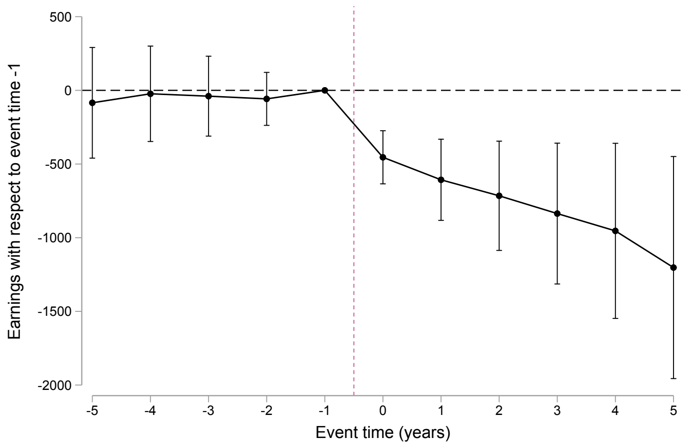
Figure 3.A.9 compares the estimated effects on maternal labor earnings with the impact on maternal post-transfer income. Both outcomes are expressed as a percentage of the mean value of the respective variable in the year prior to the shock. Compared to the impact on labor earnings, all estimates for periods after the shock are smaller (though

[22]Figure 3.A.8 shows the results for any hospital admission. I do not find that hospital admissions in general have a significant impact on maternal labor market outcomes.

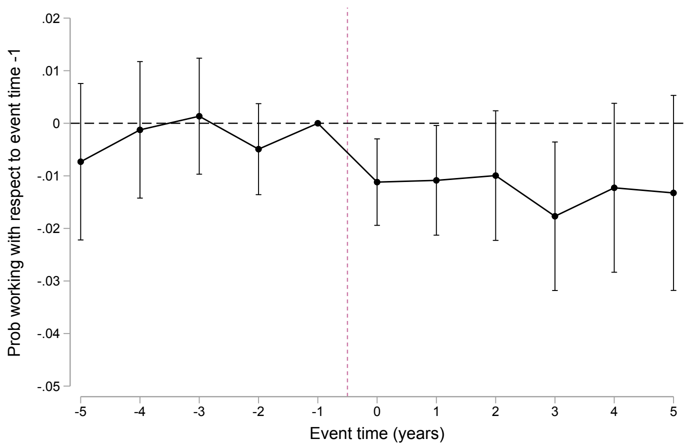
[23]Labor earnings consist of wage and salary earnings.

FIGURE 3.2: Impact of a child's severe hospitalization on maternal labor market outcomes

(a) Earnings



(b) Probability of working



Notes: This figure shows the event study graphs of the impact of a child's severe hospitalization on maternal labor market outcomes. Each figure shows the point estimates of the event time dummies with the corresponding 95 percent confidence intervals. Panel (a) plots the results for annual earnings. Panel (b) plots the results for the probability of working. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the mother level.

this difference is only significant in the period of the event). This reveals that the impact of a shock on labor earnings is partly offset through transfers. However, the drop in mothers' income still remains large: five years after the shock, their post-transfer income is around 5% lower.

Figure 3.A.10 shows the estimated coefficients for fathers and mothers as a percentage of the respective variable in the period prior to the shock. Table 3.A.2 also shows the estimated effects for fathers. Given the smaller sample, the estimates for fathers are less precise, though they do face a drop in their labor earnings. Two years after the severe hospitalization of their child, their earnings are 2.5% lower than average earnings in the year before the shock. For the two first years after the shock, the drop is similar to the estimated effect for mothers in absolute terms. However, the coefficients become relatively smaller and not significant after this period. In contrast to the result for mothers, there is no evidence of a significant drop in fathers' probability of employment. There is only a marginally significant decrease in their working probability in the year of the shock.

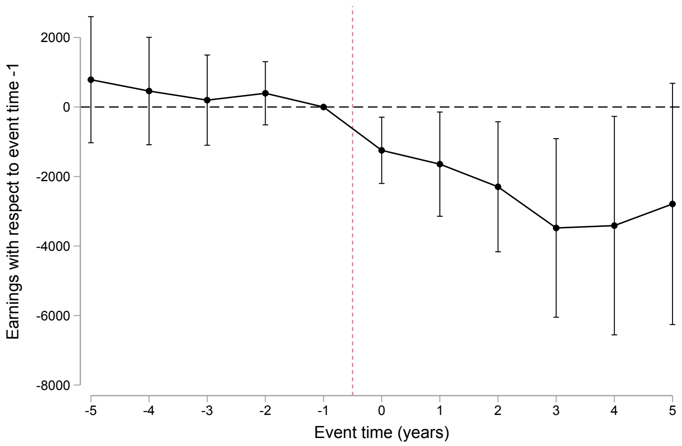
3.4.2 *Mortality*

Figure 3.3 presents the results for the impact of a child's fatal health shock on maternal earnings and labor supply. Again, there is no evidence of trends predating the event for any of the outcomes analyzed. A child's death has an enormous and long-lasting impact on maternal earnings, as shown in Panel (a). The effect is much larger than the estimated impact of a severe hospitalization. Results can be found in Table 3.A.3. Three years after the death of a child, mothers' earnings are 23% lower compared to mean earnings in the period before the event. Moreover, mothers also have a higher probability of not being employed, with a drop of 13% in their working probability three years after the event.

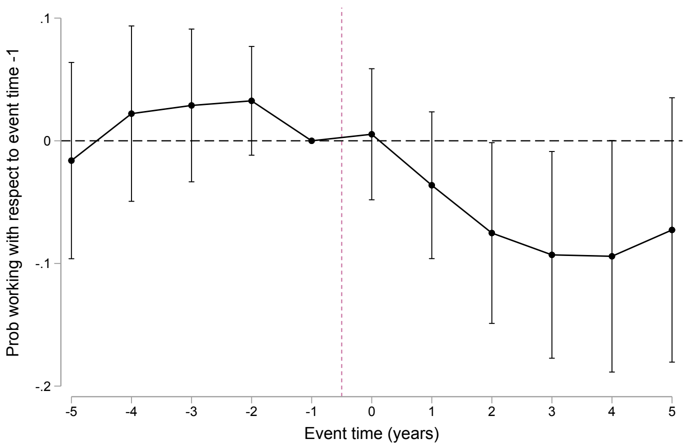
Following the death of a child, the drop in post-transfer income

FIGURE 3.3: Impact of a child's fatal health shock on maternal labor market outcomes

(a) Earnings



(b) Probability of working



Notes: This figure shows the event study graphs of the impact of a child's fatal health shock on maternal labor market outcomes. Each figure shows the point estimates of the event time dummies with the corresponding 95 percent confidence intervals. Panel (a) plots the results for annual earnings. Panel (b) plots the results for the probability of working. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the mother level.

is very similar to the estimated effect on labor earnings. Results can be found in Figure 3.A.9. This finding is consistent with the lack of bereavement support for parents who face the death of a child.

Figure 3.A.11 shows the results for fathers' and mothers' labor earnings as a percentage of mean earnings in the period prior to the event. For the first two years, the absolute drop in earnings is similar to the estimated effect for mothers. However, none of the estimates is significant (in Table 3.A.3). Fatal shocks also have a sizeable impact on fathers' probability of employment. Three years after the event, their working probability is 11% lower.

3.5 DYNAMIC DIFFERENCE-IN-DIFFERENCES

The impact of health shocks can also be examined using a simple differences-in-differences framework, by constructing counterfactuals for treated households with families who experience the shock a few years later. This quasi-experimental design exploits the potential randomness of the timing of a shock within a short period of time, a strategy that has been laid out by Fadlon and Nielsen (2017, 2019). The main difference with respect to the previous approach is that the control group is limited to households that experience the same shock at a specified later date.

Thus, the treatment group is composed of families whose child experiences the shock at a given year τ. The control group comprises households from the same age cohorts[24] whose child experienced the same shock in $\tau+\Delta$. The treatment effect is identified from the change in the difference in outcomes between the two groups over time. Crucially, there is a trade-off when choosing Δ, since a larger Δ increases the

[24]Families of the treatment and control groups are matched based on the child's and parents' years of birth. For control households, I assign a placebo "shock" at the age at which the children in the matched treatment group undergo their respective shocks.

horizon over which the effect can be observed. However, a smaller Δ is likely to capture more similar households.

In my main specification, Δ is equal to 4 years, allowing me to identify effects up to three years after the shock. After this period, the control group also undergoes a shock. In Table 3.A.4 I show that my results are robust to alternative choices of Δ.[25]

The estimating equation is a dynamic (period-by-period) difference-in-differences specification that takes the following form:

$$Y_{is} = \alpha + \beta treat_i + \sum_{t \neq -1, t=-5}^{t=3} \gamma_t \times I_t + \sum_{t \neq -1, t=-5}^{t=3} \delta_t \times I_t \times treat_i + \lambda A_{is} + \phi CBY_i + \omega_s + \varepsilon_{is} \quad (3.2)$$

Where Y_{is} denotes the outcome for parent i in calendar year s, $treat_i$ is an indicator for whether a family belongs to the treatment group, and I_t is an indicator variable for the time relative to the assigned treatment year (event time). This is the actual treatment year for the treatment group and a placebo treatment year for the control group. The parameters of interest are δ_t, which estimate the period t treatment effects relative to the period -1. I also include dummies for the age of the parent and the child's year of birth, as well as calendar-year fixed effects.

Figure 3.4a illustrates this approach for families whose children experience the shock when they are nine years old. This plot shows the raw data on maternal earnings five years before the shock and three years after the shock. The control group is made up of families whose child experienced the shock at the age of thirteen and whose family members belong to the same cohorts as the treated group. In this setting, the identifying assumption is that in the absence of the shock, the outcomes of the treatment and control groups would run parallel.

[25] Table 3.A.4 shows the results of running the same specification for different choices of Δ: Δ=3 in the first column, Δ=4 in the second column, and Δ=5 in the last column.

As shown in the graph, maternal earnings follow strikingly similar trajectories before the shock. A gap then emerges in their earnings just after the treatment group experiences the shock.

Figure 3.4b plots the coefficients and confidence intervals from the estimation of equation 3.2. There is no evidence that the trajectories of the treatment and control groups are different prior to the shock. This figure corroborates the results of the event study: the drop in maternal earnings after their child suffers a severe hospitalization is substantial and persistent. The estimated coefficients are fairly similar: three years after the shock, maternal earnings are over 1,000 lower.

3.6 HETEROGENEITY ANALYSIS

In this section, I conduct different heterogeneity analyses to shed light on the type of hospital admissions driving the impact on parental earnings. In addition, I provide further evidence to support the main results discussed in Section 3.4.1.

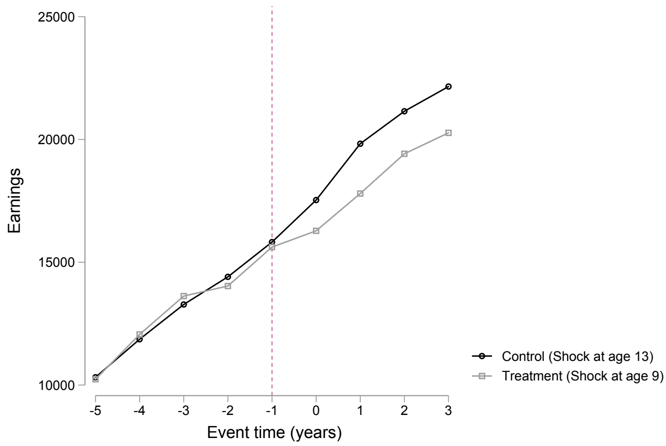
3.6.1 *Burden of Care*

If the reduction in labor earnings is partly due to the child's need for care, we would expect to find that the effect is driven by hospitalizations that impose a substantial and persistent burden of care on family members. In order to investigate this question, I empirically estimate a child's need for care one year after the shock, as measured by inpatient and outpatient visits to the hospital. I then split all hospitalizations by this measure. Figure 3.A.12 plots the average number of hospital admissions or specialist visits between one and five years after a child's hospitalization. The number of visits jumps to six visits in the year directly following the shock.

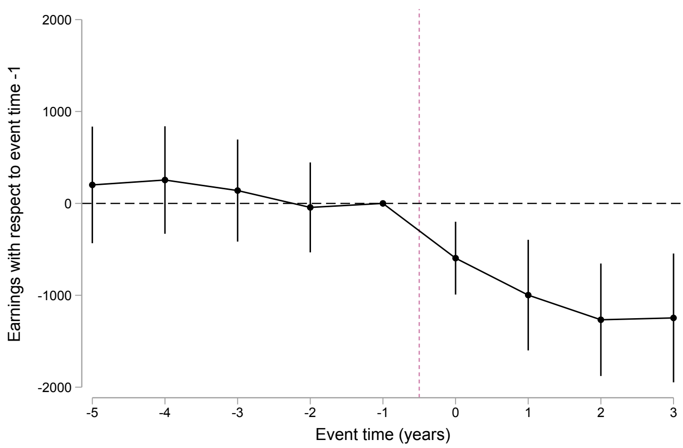
I define high burden of care hospitalizations as those requiring a number of visits one year after the shock that is greater than the av-

FIGURE 3.4: Dynamic difference-in-differences: impact of a severe hospitalization on maternal earnings

(a) Example of treatment and control group



(b) Impact of a severe hospitalization on maternal earnings



Notes: The first panel shows the raw maternal earnings data five years before and three years after the shock. The treatment group is composed of families whose child experienced the shock at age nine. The control group suffered the shock four years later. The second panel shows the coefficients and the 95 percent confidence intervals of the impact of a severe hospitalization on maternal earnings. The treatment group is composed of families whose child experiences the shock at a given year τ. The control group comprises households from the same cohorts but whose child experienced the shock at $\tau + 4$. Controls for calendar year, child's year of birth, and age of the parent are included. Standard errors are clustered at the parent level.

erage over the entire sample. Hospital admissions that require fewer visits one year after the event are defined as low burden of care hospitalizations. I apply equation 3.1 separately for these two different samples.

Figure 3.5 presents the results for maternal earnings. I do not find evidence that health shocks with a low burden of care have a significant impact on maternal earnings. In contrast, following a hospitalization that imposes a substantial burden of care, maternal earnings suffer a large and persistent decline. These results suggest that my findings are not driven by differences between families with children who suffer health shocks at different ages. Additionally, this result also suggests that the reduction in maternal labor earnings is at least partly due to the child's need for care.

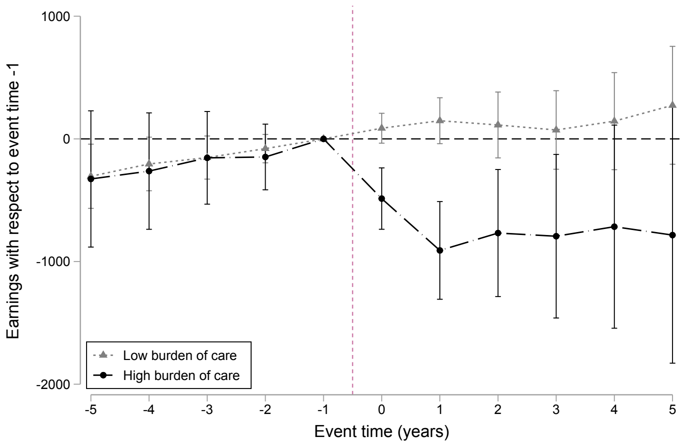
3.6.2 *Definition of Severe Hospitalizations*

Severe hospitalizations have been defined as admissions that involve a stay longer than 75% of all hospital stays. In this section, I check that my results are robust to different definitions of severe hospitalization. I then estimate the impact on maternal earnings following Equation 3.1, but for different samples of children's health shocks.

Figure 3.6 shows the results for health shocks resulting in stays longer than three days (p65), four days (p70 & p75), five days (p80), and seven days (p85). The estimated effects are robust to alternative percentile selections. All of the different definitions result in a sharp break in maternal earnings trajectories directly following the event.

Interestingly, the drop in earnings becomes more pronounced as the severity of the shock increases. Five years after a health shock resulting in a stay longer than seven days, a mother's earnings are 2,000 lower. In contrast, the maternal earnings drop following a health shock resulting in a stay of longer than three days is less than 1,000. This pattern provides further evidence that the drop is driven by health shocks

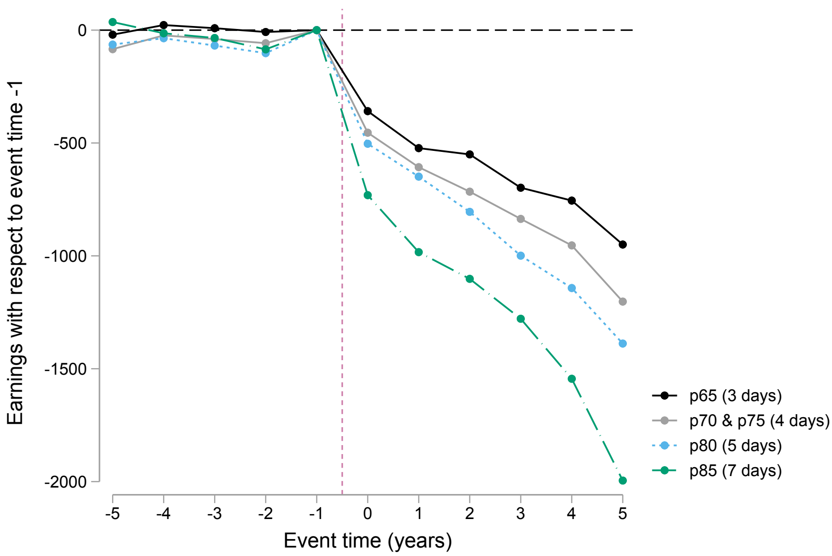
FIGURE 3.5: Impact of a child's hospitalization on maternal earnings by burden of care



Notes: This figure shows the event study graph of the impact of a child's hospitalization on maternal earnings by burden of care. The plot shows the point estimates of the event time dummies, with the corresponding 95 percent confidence intervals. Low burden of care is defined as the sample of children with a lower number of hospital visits one year after the shock than the mean for the group as a whole. High burden of care indicates the sample of children with a higher number of hospital visits one year after the shock than the mean value. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the mother level.

and not by differences due to the age at which the child experienced the shock. In the latter case, we would not expect the results to be so responsive to the degree of health shock severity.

FIGURE 3.6: Impact of a child's hospitalization on maternal earnings for different definitions of severity



Notes: This figure shows the event study graph of the impact of a child's severe hospitalization on maternal earnings for different definitions of severe hospitalization. The 65th percentile includes all hospitalizations resulting in a stay longer than three days. Percentiles 70 and 75 correspond to stays of over 4 days, while the 80th percentile involves stays longer than than five days and the 85th percentile, longer than seven days. All specifications include controls for calendar year, child's year of birth, and age of the parent.

3.6.3 *Appendicitis vs. Cancer*

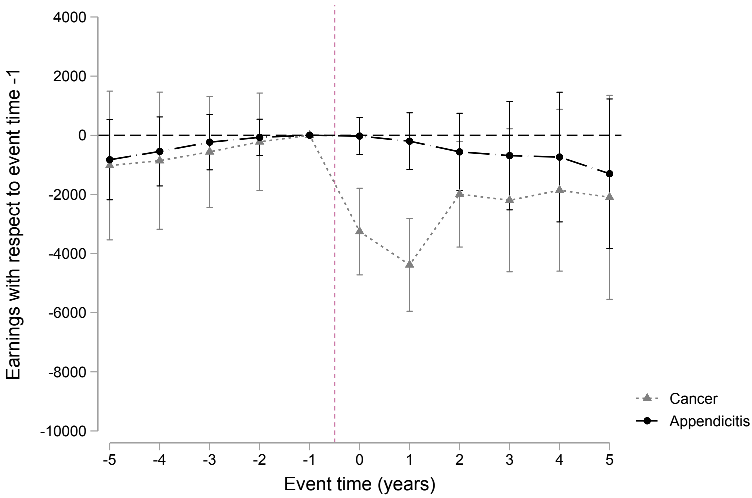
As discussed in Section 3.2, for the estimated effects on parental earnings to be caused by a child's health shock, the identifying assumption is that the child's hospitalization is uncorrelated with the counterfactual outcome, conditional on the included controls. For example, an admission caused by a deterioration of maternal earnings would violate this assumption.

I examine the validity of this identifying assumption by looking at two plausibly exogenous health shocks that are unlikely to be affected by a mother's earnings trajectory and cannot be the result of a simultaneous shock to the mother's earnings and the child's health: appendicitis and cancer. Cancer diagnoses have previously been used in the literature as exogenous health shocks (Gupta et al., 2017; Jeon and Pohl, 2017). Meanwhile, the causes and the epidemiology of appendicitis remain largely unknown (Bhangu et al., 2015; Gauderer et al., 2001).

While appendicitis is expected to generate a need for timely care, cancer is a condition with a much more complicated prognosis. In the case of cancer, involvement of family caregivers is very important in order to ensure compliance with treatments, continuity of care, and social support (Glajchen, 2004).

Figure 3.7 shows the impact of a child's hospitalization due to a diagnosis of cancer or appendicitis on maternal earnings. As expected, mothers' earnings suffer a large drop following a cancer diagnosis, while no such drop is observed following a child's hospitalization with acute appendicitis. The results of this exercise support the identifying assumption, suggesting that the observed drop in maternal earnings is not explained by mutual shocks or hospitalizations brought about due to a deterioration in maternal earnings.

FIGURE 3.7: Impact of a child's hospitalization on maternal earnings for two diagnoses: cancer and appendicitis



Notes: This figure shows the event study graph of the impact of a child's severe hospitalization on maternal earnings for two different diagnoses: cancer and appendicitis. The plot shows the point estimates of the event time dummies with the corresponding 95 percent confidence intervals. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the mother level.

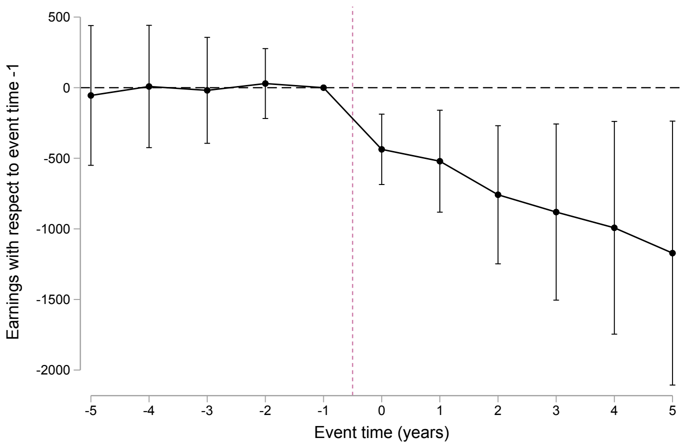
3.6.4 *Excluding Mental Health Diagnoses*

Figure 3.A.13 shows the number of severe hospital admissions broken down by diagnosis group.[26] The category[27] with the highest number of observations is mental and behavioral disorders. In order to ensure that the results are not uniquely driven by children that were admitted due to a mental health condition, I estimate Equation 3.1 once more, excluding all hospital admissions with a mental health diagnosis.

[26]Figure 3.A.14 shows the number of child fatalities broken down by cause.
[27]Classification using the chapters from the international version of the ICD-10.

The results can be found in Figure 3.8. The estimates are very similar to the main results, suggesting that the impact is not solely driven by severe hospital admissions due to mental and behavioral disorders.

FIGURE 3.8: Impact of a child's severe hospitalization on maternal earnings excluding hospitalizations with a mental health diagnosis



Notes: This figure shows the event study graph of the impact of a child's severe hospitalization on maternal earnings, excluding hospitalizations due to a mental health diagnosis. The plot shows the point estimates of the event time dummies with the corresponding 95 percent confidence intervals. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the mother level.

3.7 MECHANISMS

This section investigates potential mechanisms underpinning the observed impact of severe hospitalizations on maternal earnings. I exploit the same variation, and present the results using event studies, following the estimation of Equation 3.1.

Mental health Some studies find that parents of children with poor health or disabilities report higher stress levels and worse sleep quality (Stabile and Allin, 2012). Mental health has also been found to impact labor market outcomes (Biasi et al., 2018). In order to explore the impact of a child's severe hospitalization on parents' mental well-being, I look at the number of parental visits, with a mental health diagnosis, to a specialist or hospital.

The results presented in Figure 3.A.15 show that, with respect to the period before the shock, mothers visit specialists or hospitals at a higher rate for issues related to mental health conditions, although the effect is only significant one year after the event. The effect is much bigger for fathers, with visits increasing by nearly one and a half one year after the event. However, it becomes negative from the third year after the event. This could be driven by fathers substituting inpatient and specialist care with primary care or occupational health care doctors once they have been diagnosed.

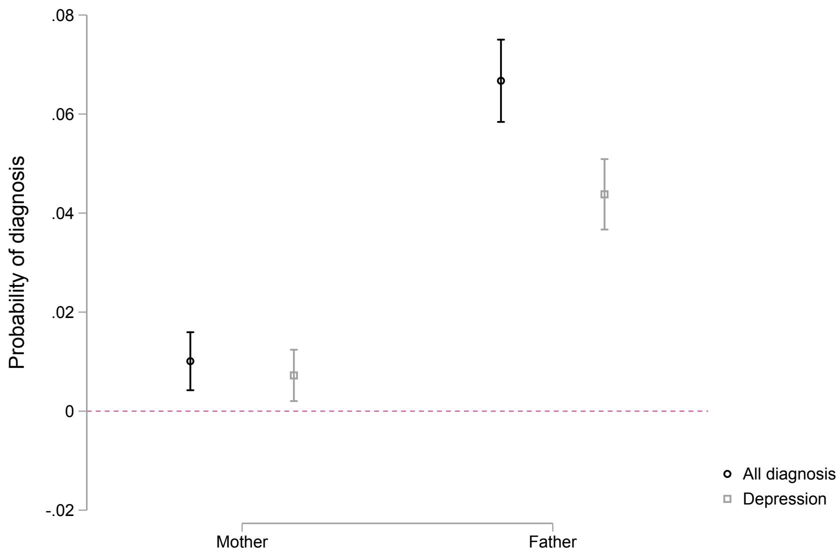
The gender differences in number of visits for mental health conditions after the shock could be explained by the data available for the analysis. The World Health Organization[28] observes that gender differences exist in patterns of help-seeking for mental health care. While women are more likely to visit a primary health care physician, men are more likely to seek a mental health specialist, and are the principal users of inpatient care.

In Figure 3.9, I plot the increase in the probability of receiving a diagnosis of a mental health condition, or a diagnosis of depression or anxiety, in the years immediately before and after the health shock. After their child's hospitalization, mothers are about 1 percentage point more likely to be diagnosed with a mental health condition, while this increases to 7 percentage points for fathers. Overall, my results suggest that parents' mental health is affected by children's health shocks. This

[28]WHO, Gender and women's mental health. https://www.who.int/mental_health/prevention/genderwomen/en/. Accessed September 2019

may, in turn, impact their labor market outcomes.

FIGURE 3.9: Impact of a child's severe hospitalization on parents' mental health



Notes: This figure shows the probability of parents receiving a diagnosis of a mental health condition, or a diagnosis of depression or anxiety, in the years immediately preceding and following their child's health shock. The plot shows the point estimates of the event time dummy for one year after the shock with the corresponding 95 percent confidence intervals. Standard errors are clustered at the parent level.

Family stability Previous papers find that having a child with a disability is associated with a higher probability of relationship dissolution (Stabile and Allin, 2012). While marital dissolution is an outcome in itself, it may also affect parents' labor supply decisions (e.g, Ananat and Michaels, 2008; Bargain et al., 2012; Leopold, 2018; Page and Stevens, 2004).

Panel (a) in Figure 3.A.17 shows the event study graph of the impact of a child's severe hospitalization on the probability of marital dissolution. I do not find evidence of an increased risk of divorce after the hospitalization of a child.

Choice of work environment Other studies have indicated that women prefer jobs that are more "family friendly" after childbirth (e.g, Goldin and Katz, 2016; Lundborg et al., 2017). In particular, Pertold-Gebicka et al. (2016) and Kleven et al. (2019b) find that mothers have a higher probability of moving to an occupation in the public sector following parenthood, which is known to have more flexible working conditions.

Similarly, after a severe hospitalization of a child, mothers may also seek a more family-friendly job in order to take care of their child. In panel (b) in Figure 3.A.17, I examine whether mothers have a higher probability of working in the public sector after their child undergoes a health shock. I do not find that this is the case, suggesting that mothers do not adjust their labor supply in this manner. More generally, panel (c) in Figure 3.A.17 looks at whether mothers have a higher probability of moving to a different job after a child's health shock. For each year, I define an indicator variable equal to one if the mother is not working in the same enterprise as in the previous period. I do not find evidence that mothers have a higher probability of switching to a different job after the health shock.

3.8 CONCLUSIONS

This paper provides new evidence on the impact of children's health shocks on parental labor market outcomes. To identify the causal effect, I compare families whose children are exposed to health shocks at varying ages, conditional on the parents' and children's ages. This allows me to abstract from differences across families who suffer the illness or death of a child and those who do not.

I use long panels of high-quality administrative data from Finland on family income and health trajectories. This enables me to exploit precisely and objectively identified health shocks and provide visually clear evidence using an event study approach. In particular, I look at the impact of severe hospitalizations, focusing on children who had not been hospitalized by school-starting age, and the impact of fatal health shocks.

The results show that children's health shocks have a detrimental and persistent impact on both parents' labor market trajectories. Three years after a severe hospitalization, mothers' earnings are 5% lower, while fathers' earnings drop by 2.5%. Additionally, I show that the impact is driven by hospitalizations that require substantial and persistent care after the event.

To put the magnitude of the effects in context, the impact on maternal earnings is approximately one fourth of the estimated effect of a health shock on an individual's own earnings (Dobkin et al., 2018; Meyer and Mok, 2019; Fadlon and Nielsen, 2017), and more than one tenth of the estimated drop in maternal earnings three years after childbirth in Finland (Sieppi and Pehkonen, 2019).

For families that face the death of a child, the impact on labor earnings is much larger: three years after the death of a child, mothers' earnings are 23% lower. For fathers, the estimated coefficients are negative and large in magnitude, but imprecise, suggesting that fatal shocks could also impact their earnings.

Children's health shocks also have an impact on parents' mental well-being, which I document using data on hospital and specialist diagnoses. The effect seems to be stronger for fathers, though this could be explained by gender differences in patterns of help-seeking for mental health issues. Assuming that the impact on earnings of a depression diagnosis is similar to the effect estimated by Biasi et al. (2018) , the increased risk of depression after a child's severe hospitalization would explain around 60% of the observed drop in earnings for fathers.

Taken together, the results point to the importance of providing assistance, and especially mental health support, to families whose child experiences a health shock. My findings also show that while the loss in earnings for parents whose child undergoes a severe hospitalization is partly offset through transfers, this is not the case for fatal shocks. This opens debate over whether the existing situation is optimal or there is room for public intervention. My study also provides useful inputs for cost-benefit studies of policies aimed at preventing children's diseases or deaths, which may wish to incorporate these indirect costs in their estimations. Finally, further research is needed in order to understand the potential spillover effects of these shocks on siblings' development and well-being.

3

APPENDIX

APPENDIX 3.A

Table 3.A.1: Summary statistics

	All hospitalizations		Severe hospitalizations		Mortality	
	mean	sd	mean	sd	mean	sd
Child characteristics						
Age at event time	10.162	3.251	11.064	3.366	11.659	4.205
Male	0.529	0.499	0.558	0.497	0.592	0.492
Mother characteristics						
Age at event time	37.046	5.760	37.772	6.101	38.260	6.162
Finnish	0.981	0.137	0.982	0.132	0.966	0.180
Single	0.035	0.185	0.047	0.211	0.036	0.186
Married	0.421	0.494	0.450	0.498	0.469	0.500
Earnings t=-1	16308	14063	16002	15055	15371	14373
Prob. working t=-1	0.740	0.439	0.713	0.452	0.729	0.445
N visits mental health t-1	0.217	2.704	0.417	3.740	0.031	0.247
Father characteristics						
Age at event time	39.619	6.278	40.534	6.574	41.064	6.457
Earnings t=-1	28670	50618	26531	24060	26826	20571
Prob. working t=-1	0.868	0.338	0.820	0.384	0.840	0.367
N visits mental health t-1	0.084	1.315	0.360	2.759	0.004	0.061
Observations	25960		8546		358	

This table shows summary statistics for the different samples included in the analysis. Columns 1 and 2 includes all children who suffered a hospitalization after age 6 (and had not been hospitalized previously), while columns 3 and 4 include all children who suffered their first severe hospitalization after age 6, and the final two columns include all children who suffered a fatal health shock after age 6.

Table 3.A.2: Impact of a child's severe hospitalization on parents' labor market outcomes

	(1) Mother earnings	(2) Father earnings	(3) Mother earnings	(4) Father earnings	(5) Mother log earnings	(6) Father log earnings	(7) Mother working	(8) Father working
Time to shock:								
-5	-84.380	273.561	-67.102	262.401	-0.042	0.034	-0.007	0.003
	(191.568)	(357.110)	(192.872)	(358.578)	(0.065)	(0.062)	(0.008)	(0.007)
-4	-23.309	281.386	-10.063	272.755	0.030	0.020	-0.001	0.002
	(165.157)	(315.166)	(165.945)	(315.569)	(0.055)	(0.051)	(0.007)	(0.006)
-3	-39.503	-26.097	-30.715	-32.045	-0.034	-0.020	0.001	-0.005
	(138.270)	(225.781)	(138.747)	(226.965)	(0.045)	(0.041)	(0.006)	(0.005)
-2	-57.869	165.399	-53.386	162.322	0.045	0.032	-0.005	0.001
	(91.723)	(166.339)	(91.640)	(167.047)	(0.032)	(0.029)	(0.004)	(0.004)
0	-454.374***	-408.386**	-458.822***	-405.127**	-0.086***	-0.051*	-0.011***	-0.006*
	(91.959)	(181.605)	(92.332)	(180.781)	(0.031)	(0.030)	(0.004)	(0.004)
1	-607.091***	-675.999**	-616.087***	-669.780**	-0.101**	-0.044	-0.011**	-0.006
	(140.516)	(264.774)	(140.698)	(263.288)	(0.043)	(0.041)	(0.005)	(0.005)
2	-715.565***	-673.930**	-729.418***	-664.589**	-0.101*	-0.064	-0.010	-0.004
	(189.240)	(339.520)	(190.436)	(337.810)	(0.052)	(0.053)	(0.006)	(0.006)
3	-836.169***	-693.665	-854.654***	-681.169	-0.158***	-0.048	-0.018**	-0.005
	(243.870)	(426.426)	(245.551)	(424.161)	(0.061)	(0.065)	(0.007)	(0.007)
4	-953.614***	-695.030	-976.50***	-679.388	-0.148**	-0.033	-0.012	-0.004
	(303.165)	(527.297)	(304.583)	(525.275)	(0.072)	(0.078)	(0.008)	(0.008)
5	-1202.544***	-787.144	-1229.833***	-768.261	-0.197**	-0.011	-0.013	-0.000
	(384.453)	(667.298)	(385.592)	(663.390)	(0.084)	(0.095)	(0.009)	(0.010)
Observations	90889	75200	90889	75200	90889	75200	90889	75200
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Additional controls	NO	NO	YES	YES	NO	NO	NO	NO
$\overline{Y}_{t-1}$	16002.080	26531.112	16002.080	26531.112	7.693	8.678	0.713	0.820

This table shows the estimates of the impact of a child's severe hospitalization on parents' labor market outcomes. All specifications include controls for calendar year, child's year of birth, and age of the parent. The first two columns show earnings in euros. Columns (3) and (4) shows the same results but controlling for the child's gender. In columns (5) and (6), earnings are expressed as logarithms. The last two columns show the probability of a parent being employed. Standard errors are clustered at the parent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.3: Impact of a child's fatal health shock on parents' labor market outcomes

	(1) Mother earnings	(2) Father earnings	(3) Mother log earnings	(4) Father log earnings	(5) Mother working	(6) Father working
Time to shock:						
-5	785.306	-743.666	-0.326	0.276	-0.016	0.035
	(925.454)	(1472.948)	(0.343)	(0.329)	(0.041)	(0.034)
-4	459.479	-467.880	-0.128	0.228	0.022	0.011
	(788.348)	(1327.571)	(0.294)	(0.272)	(0.036)	(0.030)
-3	197.646	-37.249	0.107	0.017	0.029	0.009
	(661.928)	(1003.941)	(0.240)	(0.230)	(0.032)	(0.026)
-2	394.744	372.507	0.062	0.182	0.033	0.020
	(463.736)	(668.561)	(0.176)	(0.153)	(0.023)	(0.021)
0	-1246.349**	-230.906	-0.175	-0.242	0.005	-0.005
	(486.161)	(752.163)	(0.214)	(0.174)	(0.027)	(0.021)
1	-1643.673**	-1849.579	-0.242	-0.680***	-0.036	-0.074***
	(765.151)	(1265.472)	(0.288)	(0.247)	(0.031)	(0.027)
2	-2295.207**	-2054.675	-0.590*	-1.008***	-0.075**	-0.075**
	(954.504)	(1687.604)	(0.352)	(0.316)	(0.038)	(0.035)
3	-3479.772***	-2187.651	-0.900**	-1.109***	-0.093**	-0.088**
	(1310.418)	(2221.702)	(0.407)	(0.395)	(0.043)	(0.041)
4	-3412.421**	-1665.007	-0.924*	-1.157**	-0.094*	-0.083*
	(1603.756)	(2817.436)	(0.473)	(0.480)	(0.048)	(0.048)
5	-2789.245	-1543.083	-0.911*	-1.270**	-0.073	-0.103*
	(1770.657)	(3163.789)	(0.548)	(0.581)	(0.055)	(0.057)
Observations	3126	2743	3126	2743	3126	2743
Controls	YES	YES	YES	YES	YES	YES
$\overline{Y}_{t-1}$	14744.982	25069.911	7.418	8.100	0.707	0.757

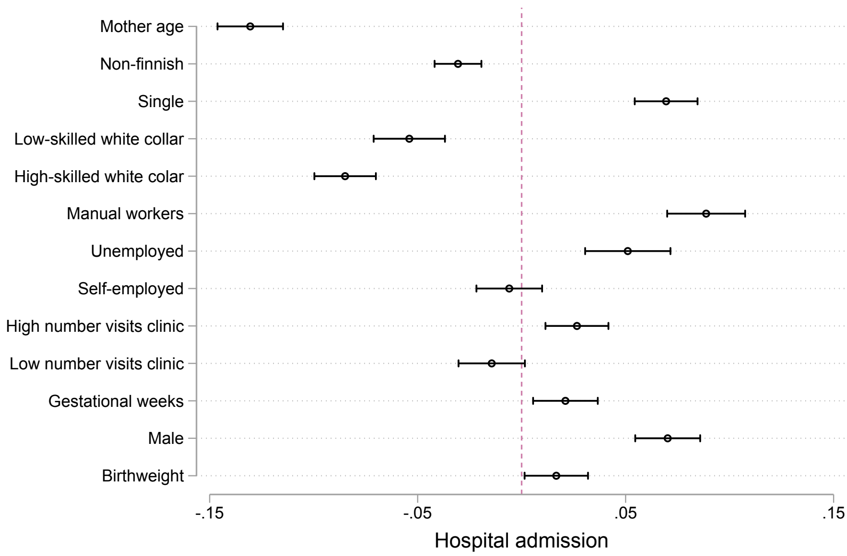
This table shows the estimates of the impact of a fatal health shock on parents' earnings, expressed in euros and as a logarithm, and working probability. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the parent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.4: Dynamic differences-in-differences: Choice of Δ

	(1) $\Delta = 3$	(2) $\Delta = 4$	(3) $\Delta = 5$
event time -5* treat	60.822	201.141	283.636
	(313.267)	(323.677)	(343.432)
event time -4* treat	77.134	254.501	454.932
	(320.463)	(298.332)	(326.045)
event time -3* treat	-53.353	139.822	120.111
	(290.405)	(283.309)	(322.221)
event time -2* treat	-59.654	-43.648	112.606
	(207.239)	(249.614)	(234.165)
event time 0* treat	-582.327**	-595.670***	-526.590***
	(269.223)	(202.105)	(256.658)
event time 1* treat	-496.220*	-997.970***	-742.260***
	(296.181)	(306.989)	(301.296)
event time 2* treat	-1107.263***	-1266.167***	-1103.960***
	(389.381)	(312.216)	(315.412)
event time 3* treat		-1246.082***	-988.627***
		(357.562)	(369.392)
event time 4* treat			-1114.424***
			(406.475)
Observations	53948	61341	69049
Controls	YES	YES	YES

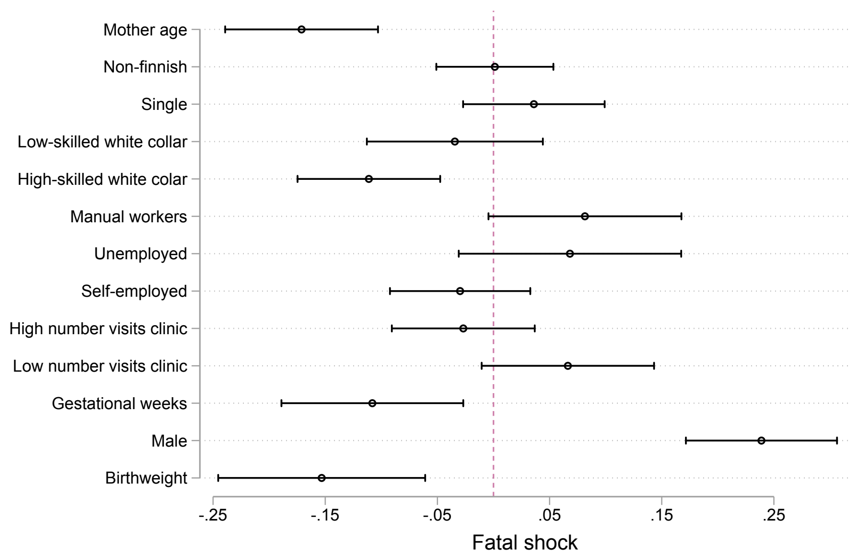
This table shows the estimates of the impact of a severe hospital admission on maternal labor earnings. The treatment group is composed of families whose child experience the shock at a given year τ. The control group comprises households from the same parental and child age cohorts whose child experienced the shock in $\tau + \Delta$. Each column shows the results for different selections of Δ. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the parent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

FIGURE 3.A.1: Differences in characteristics: across families



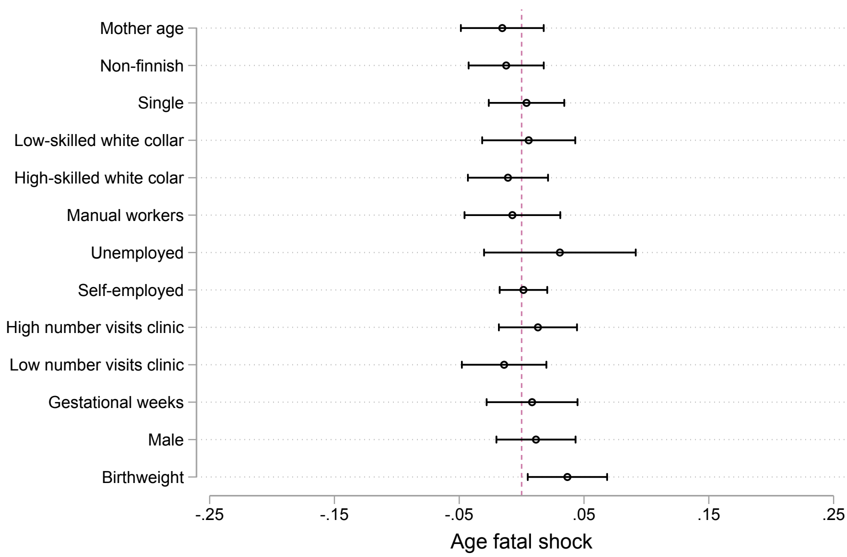
Notes: The figure shows the coefficients and 95% CI from separate regressions of each (standardized) variable on an indicator that takes a value of 1 if the child suffered a hospitalization. All specifications include year-of-birth fixed effects. Standard errors are clustered at the mother level.

FIGURE 3.A.2: Differences in characteristics (mortality sample): across families



Notes: The figure shows the coefficients and 95% CI from separate regressions of each (standardized) variable on an indicator that takes a value of 1 if the child suffered a fatal shock. All specifications include year-of-birth fixed effects. Standard errors are clustered at the mother level.

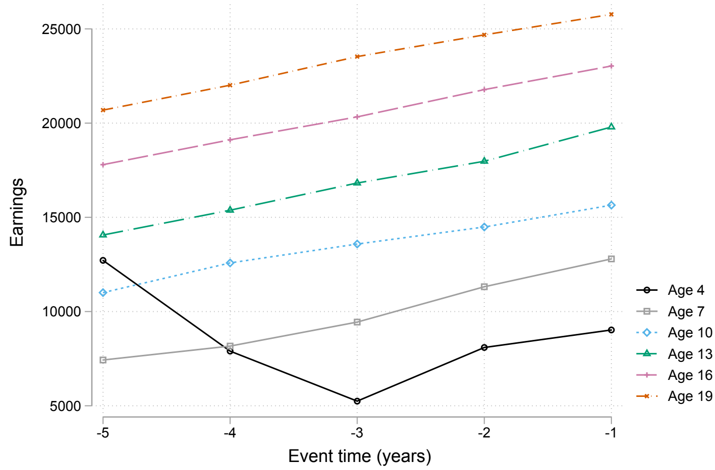
FIGURE 3.A.3: Differences in characteristics (mortality sample): within affected families



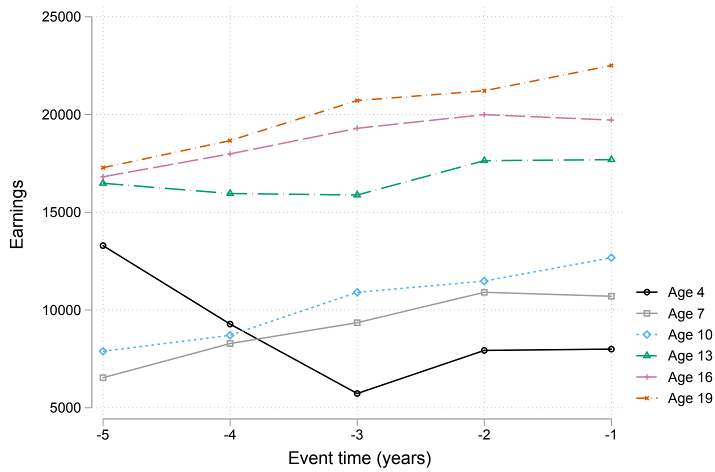
Notes: The figure shows the coefficients and 95% CI from separate regressions of each (standardized) variable on children's age at death. All specifications include year-of-birth fixed effects. Standard errors are clustered at the mother level.

FIGURE 3.A.4: Raw maternal earnings trajectories before the event by children's age at hospital admission

(a) Severe hospitalizations: by age at hospital admission

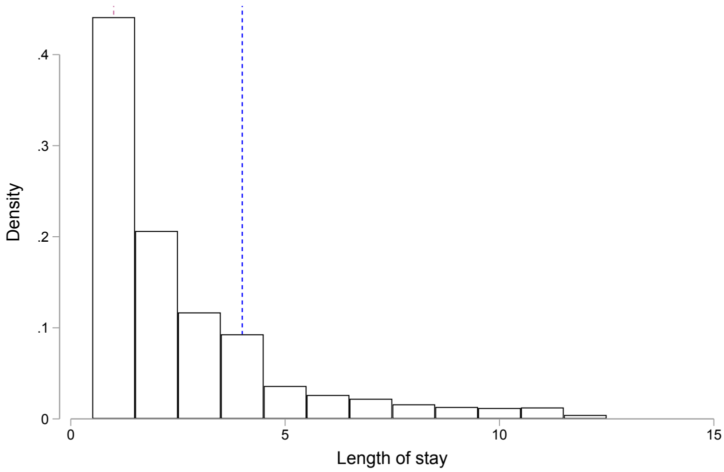


(b) Mortality: by age



Notes: This figure shows the raw maternal earnings trajectories by event time for each age group. Panel (a) shows the yearly average earnings for the years leading up to a severe hospitalization, by the children's age at admission. Panel (b) shows the analogous graph but for mortality.

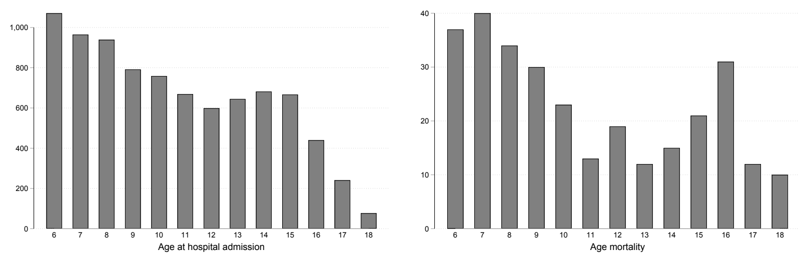
FIGURE 3.A.5: Distribution of length of hospitalization



Notes: This figure shows the distribution of the length of hospitalizations. The blue dashed line shows the 75th percentile value, which corresponds to four days.

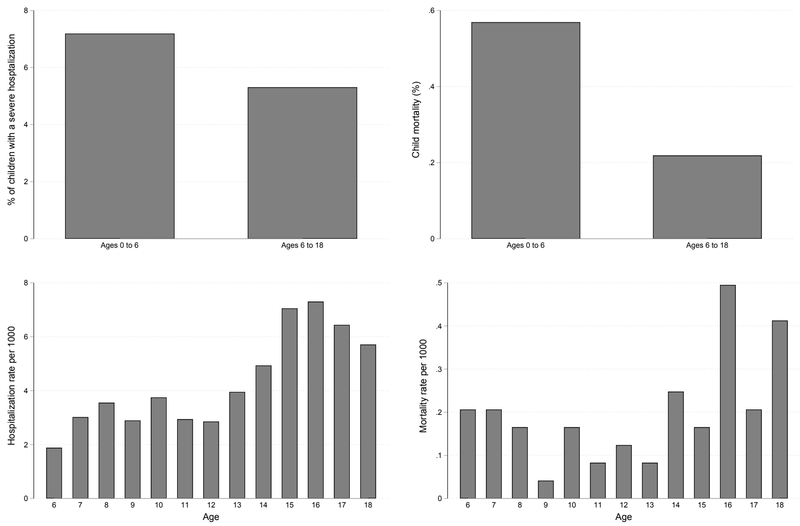
FIGURE 3.A.6: Number of observations by age at event time

(a) Severe hospitalizations (b) Mortality



Notes: This figure shows the number of observations by the age of the child at hospital admission, in panel (a), and the number of observations by the age of the child at the time of the fatal shock, panel (b).

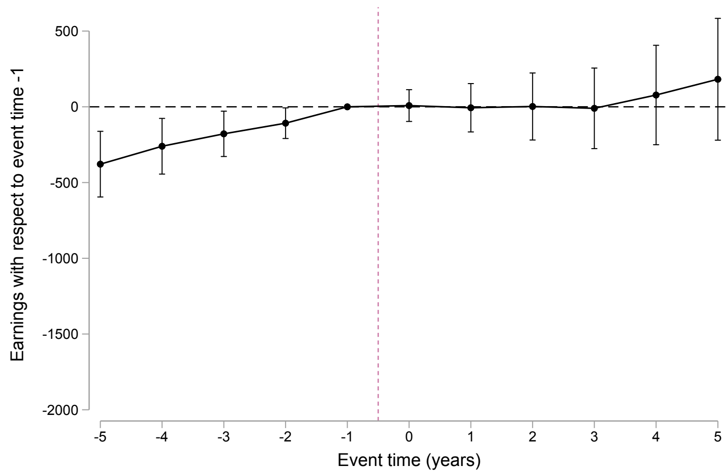
FIGURE 3.A.7: Descriptive: children born in 1990



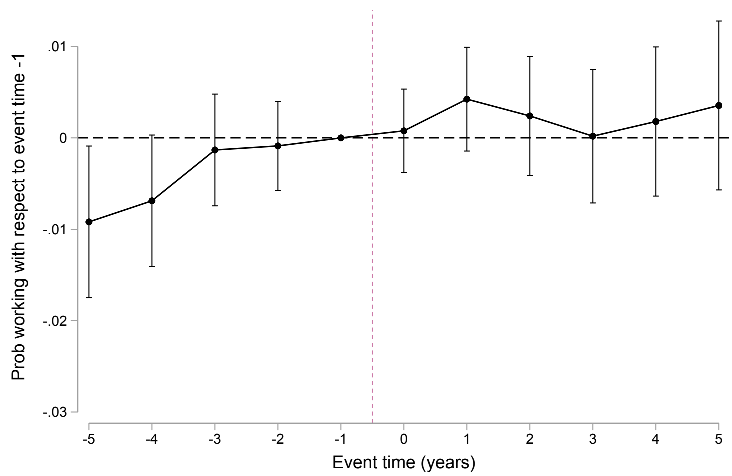
Notes: This figure provides different descriptive graphs for the sample of children born in 1990. The upper-left figure shows the percentage of children in two age groups (0-6 and 6-18) that suffered a severe hospitalization. The upper-right figure shows child mortality for the same age groups. The figure in the lower left shows the severe hospitalization rate per 1000 children by age, while the figure in the lower right shows the same figures for child mortality.

FIGURE 3.A.8: Impact of any hospital admission on maternal labor market outcomes

(a) Earnings



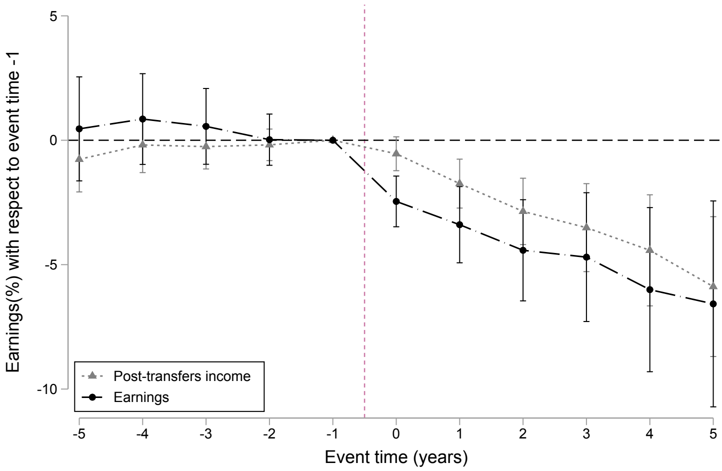
(b) Probability of working



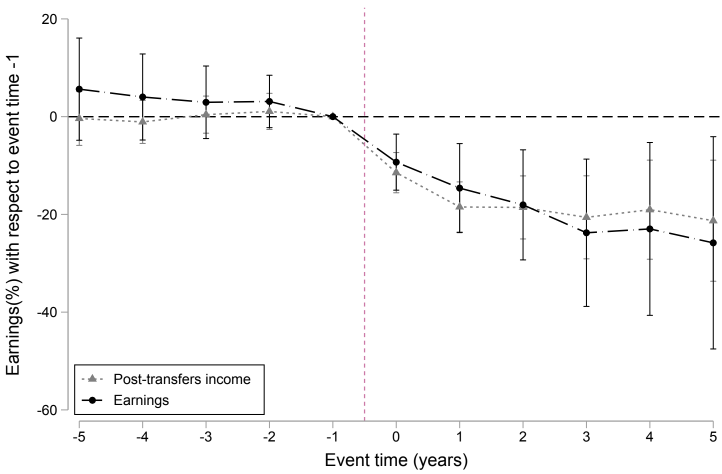
Notes: This figure shows the event study graphs of the impact of hospitalization of a child of any duration on maternal labor market outcomes. The plot shows the point estimates of the event time dummies with the corresponding 95 percent confidence intervals. Panel (a) plots the coefficients on earnings for mothers and panel (b) plots the coefficients on the probability of working. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the parent level.

FIGURE 3.A.9: Impact of a child's health shock on parents' post-transfers income

(a) Severe hospital admissions

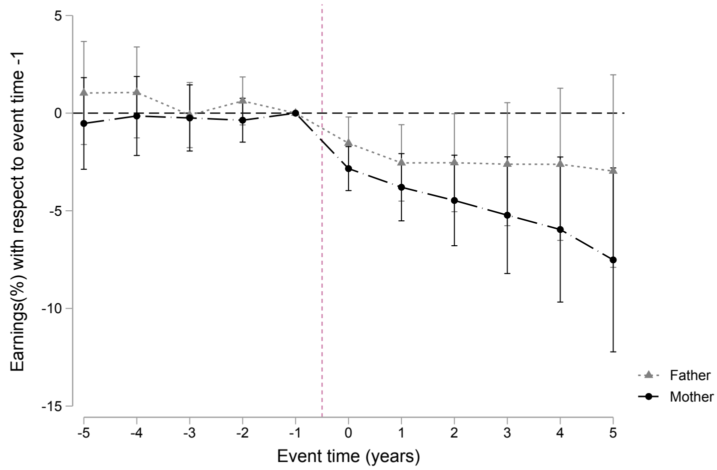


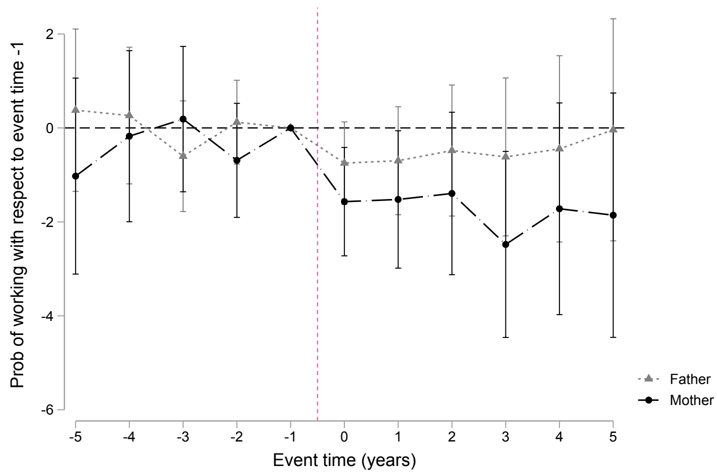
(b) Fatal shocks



Notes: This figure shows the event study graphs of the impact of a child's severe hospitalization on post-transfer maternal income (grey line) and labor earnings (black line) for severe hospitalizations (panel (a)) and for mortality (panel (b)). Each figure shows the point estimates of the event time dummies as a percentage of the period prior to the shock with the corresponding 95 percent confidence intervals. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the parent level.

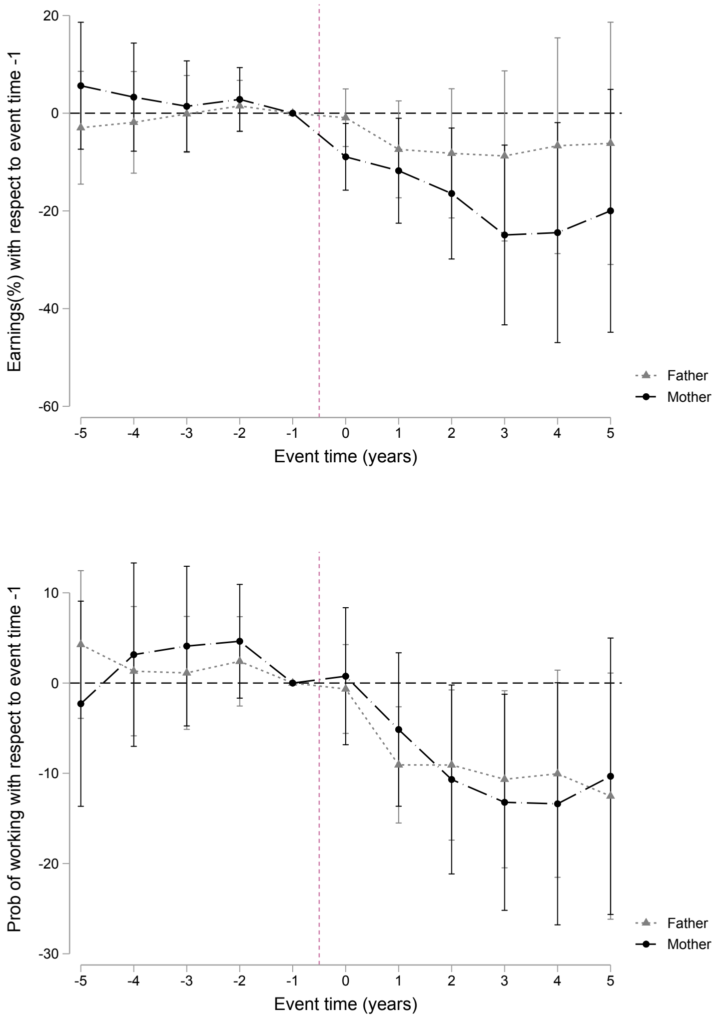
FIGURE 3.A.10: Impact of a child's severe hospitalization on parental labor market outcomes





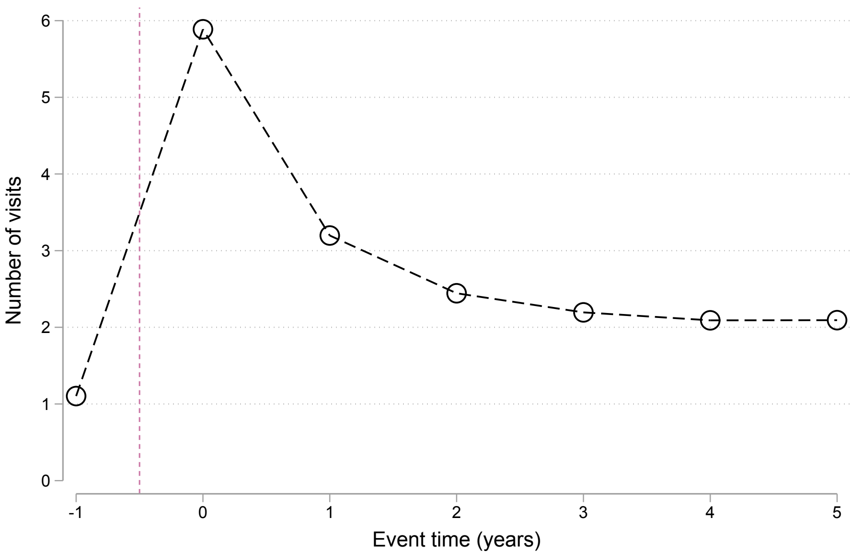
Notes: This figure shows the event study graphs of the impact of a child's severe hospitalization on labor market outcomes for both parents. Each figure shows the point estimates of the event time dummies, as a percentage of the period prior to the shock, with the corresponding 95 percent confidence intervals. Panel (a) plots the results for annual earnings. Panel (b) plots the results for the probability of working. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the parent level.

FIGURE 3.A.11: Impact of a child's fatal health shock on parental labor market outcomes



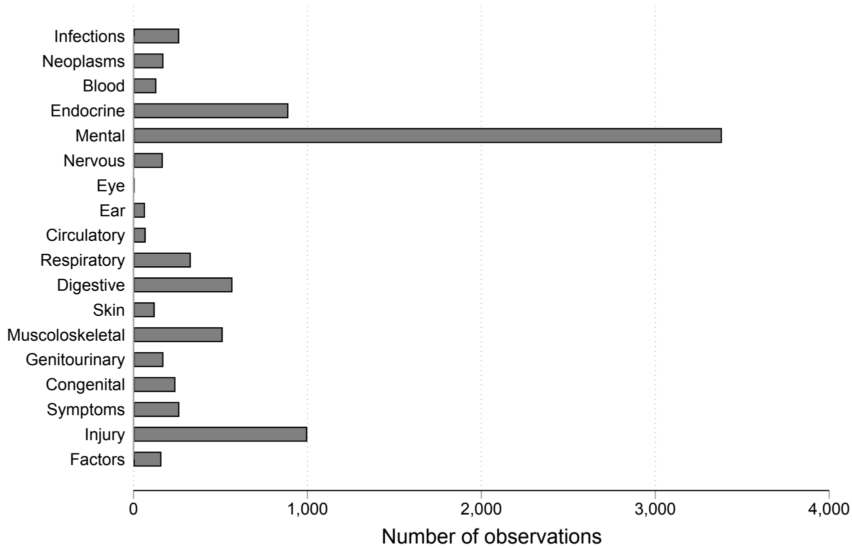
Notes: This figure shows the event study graphs of the impact of a child's death on labor market outcomes for both parents. Each figure shows the point estimates of the event time dummies, as a percentage of the period prior to the shock, with the corresponding 95 percent confidence intervals. Panel (a) shows the results for earnings and panel (b) for the probability of working. All specifications include controls for calendar year, child's year of birth, and age of the parent. Standard errors are clustered at the parent level.

FIGURE 3.A.12: Children's number of visits before and after a hospital admission



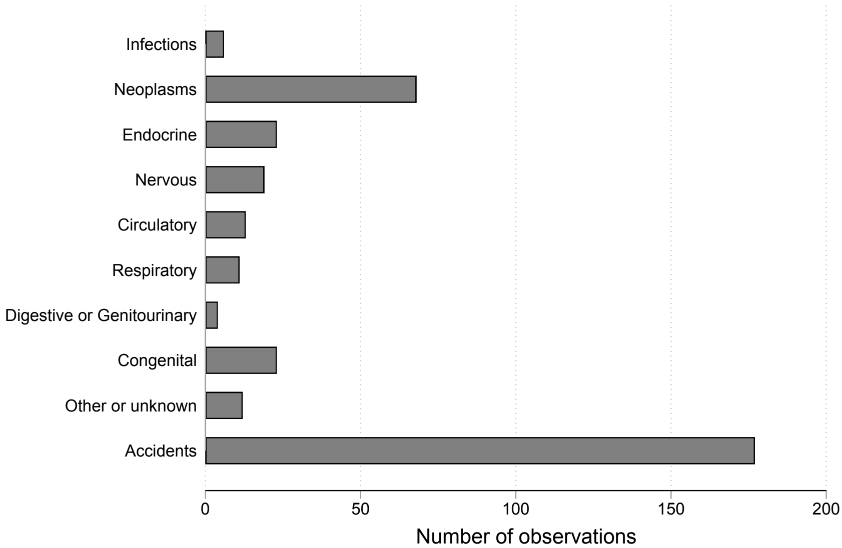
Notes: The figure shows children's average number of specialist or hospital visits one year before and five years after a hospital admission.

FIGURE 3.A.13: Severe hospitalizations: by main diagnosis group



Notes: This figure shows the number of children who suffered a severe hospitalization by main diagnosis group (ICD-10 Chapters). Categories include: Certain infectious and parasitic diseases, Neoplasms, Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism, Endocrine, nutritional and metabolic diseases, Mental and behavioural disorders, Diseases of the nervous system, Diseases of the eye and adnexa, Diseases of the ear and mastoid process, Diseases of the circulatory system, Diseases of the respiratory system, Diseases of the digestive system, Diseases of the skin and subcutaneous tissue,Diseases of the musculoskeletal system and connective tissue, Diseases of the genitourinary system, Congenital malformations, deformations and chromosomal abnormalities, Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified, Injury, poisoning and certain other consequences of external causes, and Factors influencing health status and contact with health services.

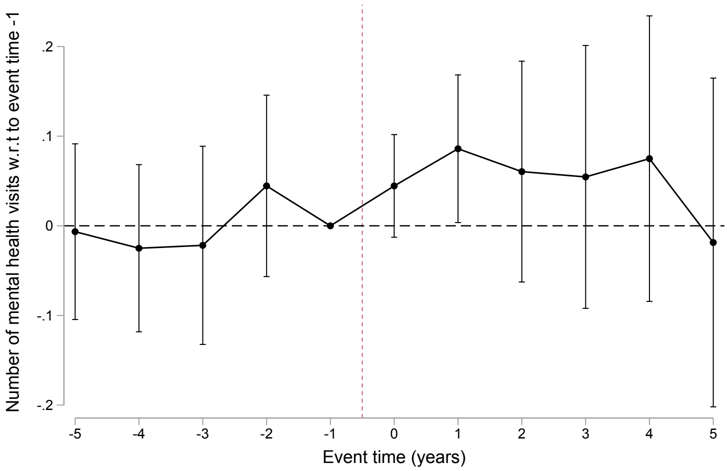
FIGURE 3.A.14: Mortality: by main cause



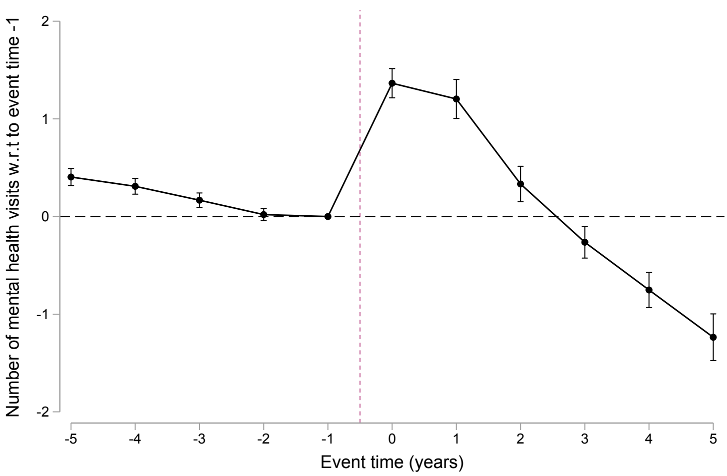
Notes: This figure shows the number of children who suffered a fatal health shocks by main cause of death. Categories include: Certain infectious and parasitic diseases, Neoplasms, Endocrine, nutritional and metabolic diseases, Diseases of the circulatory system excl. alcohol-related, Diseases of the respiratory system, Diseases of the digestive system, Diseases of the genitourinary system, Congenital malformations, Other diseases excl. alcohol-related, Ill-defined and unknown causes of mortality, Alcohol-related diseases and accidental poisoning by alcohol, Accidents and violence excl. accidental poisoning by alcohol.

FIGURE 3.A.15: Impact of a child severe hospital admission on parents mental health

(a) Mother



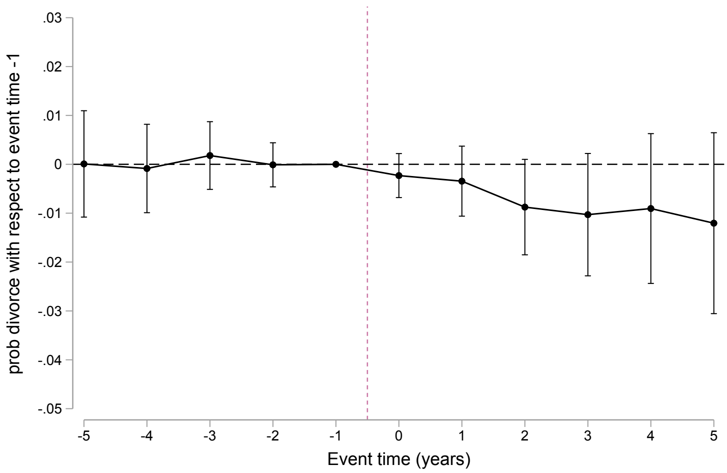
(b) Father



Notes: This figure shows the number of children who suffered a fatal health shocks by main cause of death. Categories include: Certain infectious and parasitic diseases, Neoplasms, Endocrine, nutritional and metabolic diseases, Diseases of the circulatory system excl. alcohol-related, Diseases of the respiratory system, Diseases of the digestive system, Diseases of the genitourinary system, Congenital malformations, Other diseases excl. alcohol-related, Ill-defined and unknown causes of mortality, Alcohol-related diseases and accidental poisoning by alcohol, Accidents and violence excl. accidental poisoning by alcohol.

FIGURE 3.A.16: Mechanisms: impact of a child severe hospital admission on family stability

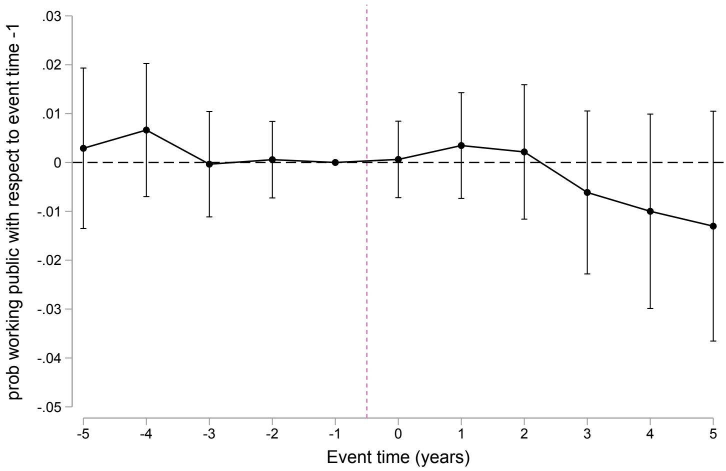
(a) Figures/Probability of divorce



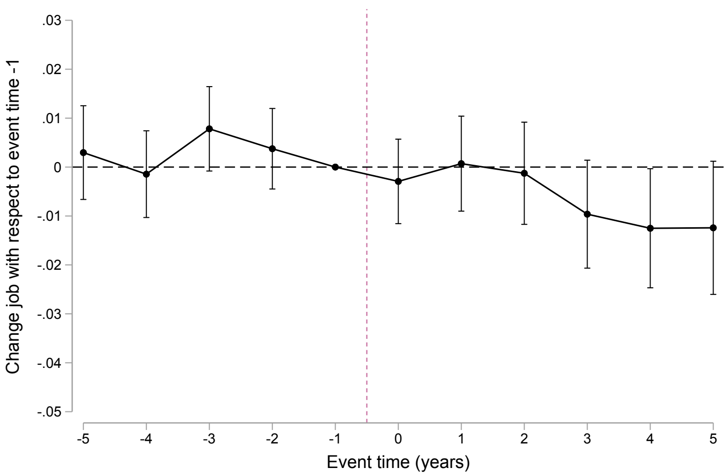
Notes: This figure shows the event study graphs of the impact of a child severe hospital admission on the probability of relationship dissolution. The figure shows the point estimates of the event time dummies, with the corresponding 95 percent confidence intervals. $\overline{Y}_{t-1}$ is 0.185 for probability of divorce. Controls for calendar year, child year of birth and age of the parent are included. Standard errors are clustered at the parent level.

FIGURE 3.A.17: Mechanisms: impact of a child severe hospital admission on choice of working environment

(a) Probability of working in the public sector



(b) Probability of switching jobs



Notes: This figure shows the event study graphs of the impact of a child severe hospital admission on the probability of working in the public sector (panel (a)) and the probability of switching jobs (panel (b)). Each figure shows the point estimates of the event time dummies, with the corresponding 95 percent confidence intervals. $\overline{Y}_{t-1}$ is 0.331 for probability of working in a state-owned enterprise and 0.156 for probability of switching jobs. All specifications include controls for calendar year, child year of birth and age of the parent. Standard errors are clustered at the parent level.

BIBLIOGRAPHY

Adda, J., Dustmann, C., and Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, 125(2):293–337.

Adecco Healthcare (2017). Empleo en el sector sanitario. Adecco Group. https://adecco.es/wp-content/uploads/2017/07/NdP-Las-ofertas-de-empleo-en-el-sector-sanitario-se-incrementarán-un-40-según-Adecco-Healthcare.pdf. Accessed: 26-12-2017.

AEPED (2014). Test de apgar. Asociación Española de Pediatría (Spanish Pediatrics Association). http://www.aeped.es/rss/en-familia/test-apgar. Accessed: 20-12-2017.

Aizer, A., Stroud, L., and Buka, S. (2016). Maternal Stress and Child Outcomes: Evidence from Siblings. *Journal of Human Resources*, 51(3):523–555.

Almond, D., Chay, K. Y., and Lee, D. S. (2005). The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3):1031–1083.

Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25(3):153–72.

Almond, D., Currie, J., and Duque, V. (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature*, 56:1360–1446.

Almond, D., Edlund, L., and Palme, M. (2009). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly Journal of Economics*, 124(4):1729–1772.

Almqvist, C., Cnattingius, S., Lichtenstein, P., and Lundholm, C. (2012). The impact of birth mode of delivery on childhood asthma and allergic diseases–a sibling study. *Clinical & Experimental Allergy*, 42(9):1369–1376.

American College of Obstetricians and Gynecologists (2006). Mode of Term Singleton Breech Delivery. ACOG Commitee Opinion No. 340. *Obstetrics and Gynecology*, 108:235–237.

American College of Obstetricians and Gynecologists (2010). ACOG Practice bulletin no. 115: Vaginal birth after previous cesarean delivery. *Obstetrics and Gynecology*, 116(2):450–63.

American College of Obstetricians and Gynecologists (2015). Practice Bulletin No. 154: Operative Vaginal Delivery. *Obstetrics and Gynecology*, 126(5):e56–e65.

Ananat, E. O. and Michaels, G. (2008). The effect of marital breakup on the income distribution of women with children. *Journal of Human Resources*, 43(3):611–629.

Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, 34(3):545–579.

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88(3):450–477.

APA (2017). Cesarean procedure. American Pregnancy Association. http://americanpregnancy.org/labor-and-birth/cesarean-procedure/. Accessed: 11-12-2017.

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53(5):941–973.

Asher, I. and Pearce, N. (2014). Global burden of asthma among children. *The International Journal of Tuberculosis and Lung Disease*, 18(11):1269–1278.

Bager, P., Wohlfahrt, J., and Westergaard, T. (2008). Caesarean delivery and risk of atopy and allergic disesase: Meta-analyses. *Clinical & Experimental Allergy*, 38(4):634–642.

Baicker, K., Buckles, K. S., and Chandra, A. (2006). Geographic variation in the appropriate use of cesarean delivery. *Health Affairs*, 25(5):w355–w367.

Barber, E. L., Lundsberg, L. S., Belanger, K., Pettker, C. M., Funai, E. F., and Illuzzi, J. L. (2011). Indications Contributing to the Increasing Cesarean Delivery Rate. *Obstetrics and Gynecology*, 118(1):29–38.

Bargain, O., González, L., Keane, C., and Özcan, B. (2012). Female labor supply and divorce: New evidence from ireland. *European Economic Review*, 56(8):1675 – 1691.

Baydar, N., Joesch, J. M., Kieckhefer, G., Kim, H., and Greek, A. (2007). Employment behaviors of mothers who have a child with asthma. *Journal of Family and Economic Issues*, 28(3):337–355.

Benard, S., Correll, S., and Paik, I. (2007). Getting a job: Is there a motherhood penalty? Technical report.

Bentley, J. P., Roberts, C. L., Bowen, J. R., Martin, A. J., Morris, J. M., and Nassar, N. (2016). Planned Birth Before 39 Weeks and Child Development: A Population-Based Study. *Pediatrics*, 138(6):e20162002–e20162002.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–55.

Bhangu, A., Søreide, K., Di Saverio, S., Assarsson, J. H., and Drake, F. T. (2015). Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *The Lancet*, 386(10000):1278–1287.

Bhattacharya, J., Goldman, D., and McCaffrey, D. (2006). Estimating probit models with self-selected treatments. *Statistics in Medicine*, 25(3):389–413.

Biasi, B., S.Dahl, M., and Moser, P. (2018). Carrer effects of mental health. Available at SSRN: https://ssrn.com/abstract=2544251.

Black, S. E., Breining, S., Figlio, D. N., Guryan, J., Karbownik, K., Nielsen, H. S., Roth, J., and Simonsen, M. (2017). Sibling spillovers. Working Paper 23062, National Bureau of Economic Research.

Blustein, J. and Liu, J. (2015). Time to consider the risks of caesarean delivery for long term child health. *BMJ*, 350.

Boerma, T., Ronsmans, C., Melesse, D. Y., Barros, A. J., Barros, F. C., Juan, L., Moller, A. B., Say, L., Hosseinpoor, A. R., Yi, M., de Lyra Rabello Neto, D., and Temmerman, M. (2018). Global epidemiology of use of and disparities in caesarean sections. *The Lancet*, 392(10155):1341–1348.

Boksa, P. and Zhang, Y. (2008). Epinephrine administration at birth prevents long-term changes in dopaminergic parameters caused by cesarean section birth in the rat. *Psychopharmacology*, 200(3):381–391.

Bound, J., Schoenbaum, M., Stinebrickner, T. R., and Waidmann, T. (1999). The dynamic effects of health on the labor force transitions of older workers. *Labour Economics*, 6(2):179 – 202.

Bronars, S. G. and Grogger, J. (1994). The economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review*, 84(5):1141–1156.

Brown, H. (1996). Physician demand for leisure: implications for cesarean section rates. *Journal of Health Economics*, 15(2):233 – 242.

Burton, P., Chen, K., Lethbridge, L., and Phipps, S. (2017). Child health and parental paid work. *Review of Economics of the Household*, 15(2):597–620.

Bütikofer, A., Jensen, S., and Salvanes, K. G. (2018). The role of parenthood on the gender gap among top earners. *European Economic Review*, 109:103 – 123. Gender Differences in the Labor Market.

Cai, L., Mavromaras, K., and Oguzoglu, U. (2014). The effects of health status and health shocks on hours worked. *Health Economics*, 23(5):516–528.

Card, D., Fenizia, A., and Silver, D. (2019). The impacts of hospital delivery practices on infant health. *NBER Working Paper, 25986*.

Cardwell, C. R., Stene, L. C., Joner, G., Cinek, O., Svensson, J., Goldacre, M. J., Parslow, R. C., Pozzilli, P., Brigis, G., Stoyanov, D., Urbonaité, B., Šipetić, S., Schober, E., Ionescu-Tirgoviste, C., Devoti, G., de Beaufort, C. E., Buschard, K., and Patterson, C. C. (2008). Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. *Diabetologia*, 51(5):726–735.

Chiburis, R. C., Das, J., and Lokshin, M. (2012). A practical comparison of the bivariate probit and linear iv estimators. *Economics Letters*, 117(3):762 – 766.

Collado, M. C., Rautava, S., Isolauri, E., and Salminen, S. (2015). Gut microbiota: a source of novel tools to reduce the risk of human disease? *Pediatric Research*, 77:182 – 188. Review.

Costa-Ramón, A. M., Rodríguez-González, A., Serra-Burriel, M., and Campillo-Artero, C. (2018). It's about time: Cesarean sections and neonatal health. *Journal of Health Economics*, 59:46 – 59.

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

Currie, J. and Macleod, W. B. (2017). Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians. *Journal of Labor Economics*, 35(1):1–43.

Davidson, R., Roberts, S. E., Wotton, C. J., and Goldacre, M. J. (2010). Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: evidence from the Oxford record linkage study. *BMC Pulmonary Medicine*, 10(1):14.

Deaton, A. (2013). *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton University Press.

Dobkin, C., Finkelstein, A., Kluender, R., and Notowidigdo, M. J. (2018). The Economic Consequences of Hospital Admissions. *American Economic Review*, 108(2):308–352.

Dominguez-Bello, M. G., De Jesus-Laboy, K. M., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., Bokulich, N. A., Song, S. J., Hoashi, M., Rivera-Vinas, J. I., Mendez, K., Knight, R., and Clemente, J. C. (2016). Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature Medicine*, 22:250–253.

Drucker, A. M., Wang, A. R., Li, W. Q., Sevetson, E., Block, J. K., and Qureshi, A. A. (2017). The Burden of Atopic Dermatitis: Summary of a Report for the National Eczema Association. *Journal of Investigative Dermatology*, 137(1):26–30.

Evans, M. I., Richardson, D. A., Sholl, J. S., and Johnson, B. A. (1984). Cesarean section. Assessment of the convenience factor. *J Reprod Med*, 29(9):670–676.

Fadlon, I. and Nielsen, T. H. (2017). Family Labor Supply Responses to Severe Health Shocks.

Fadlon, I. and Nielsen, T. H. (2019). Family health behaviors. *American Economic Review*, 109(9):3162–91.

Fernández-Kranz, Lacuesta, A., and Rodríguez-Planas, N. (2013). The Motherhood Earnings Dip: Evidence from Administrative Records. *Journal of Human Resources*, 48(1):169–197.

Figlio, D., Guryan, J., Karbownik, K., and Roth, J. (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review*, 104(12):3921–55.

Finkelstein, E. A., Graham, W. C. K., and Malhotra, R. (2014). Lifetime Direct Medical Costs of Childhood Obesity. *Pediatrics*, 133(5):854–862.

Finnish National Agency for Education (2018). Compulsory education in finland. https://www.oph.fi/en/statistics-and-publications/publications/compulsory-education-finland Accessed: 13-08-2019.

Fraser, W., Usher, R. H., McLean, F. H., Bossenberry, C., Thomson, M. E., Kramer, M. S., Smith, L., and Power, H. (1987). Temporal variation in rates of cesarean section for dystocia: Does "convenience" play a role? *American Journal of Obstetrics and Gynecology*, 156(2):300 – 304.

García-Gómez, P. (2011). Institutions, health shocks and labour market outcomes across europe. *Journal of Health Economics*, 30(1):200 – 213.

García-Gómez, P., van Kippersluis, H., O'Donnell, O., and van Doorslaer, E. (2013). Long-term and spillover effects of health shocks on employment and income. *Journal of Human Resources*, 48(4):873–909.

Gauderer, M. W., Crane, M. M., Green, J. A., DeCou, J. M., and Abrams, R. S. (2001). Acute appendicitis in children: The importance of family history. *Journal of Pediatric Surgery*, 36(8):1214 – 1217.

Glajchen, M. (2004). The emerging role and needs of family caregivers in cancer care. *J Support Oncol*, 2(2):145–155.

Global Asthma Network (2018). *The Global Asthma Report 2018*. Global Asthma Network, Auckland, New Zealand.

Goldin, C. and Katz, L. F. (2016). A most egalitarian profession: Pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics*, 34(3):705–746.

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

Grivell, R. M. and Dodd, J. M. (2011). Short- and long-term outcomes after cesarean section. *Expert Review of Obstetrics and Gynecology*, 6(2):205–215.

Gruber, J. and Owings, M. (1996). Physician Financial Incentives and Cesarean Section Delivery. *The RAND Journal of Economics*, 27(1):99.

Gupta, A., Morrison, E. R., Fedorenko, C., and Ramsey, S. (2017). Leverage, default, and mortality: Evidence from cancer diagnoses. *Columbia Law and Economics Working Paper*, (514):15–35.

Halla, M., Mayr, H., Pruckner, G. J., and Garcia-Gomez, P. (2016). Cutting fertility? the effect of cesarean deliveries on subsequent fertility and maternal labor supply. IZA Discussion Paper.

Hannah, M. (1989). Birth asphyxia: does the Apgar score have diagnostic value? *Obstetrics and Gynecology*, 73(2):299—300.

Hannah, M. E., Hannah, W. J., Hewson, S. A., Hodnett, E. D., Saigal, S., Willan, A. R., and Collaborative, T. B. T. (2000). Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. *The Lancet*, 356(9239):1375–1383.

Hansen, A. K., Wisborg, K., Uldbjerg, N., and Henriksen, T. B. (2008). Risk of respiratory morbidity in term infants delivered by elective caesarean section: cohort study. *BMJ : British Medical Journal*, 336:85–87.

Harjutsalo, V. (2008). Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *Lancet (London, England)*, 371(9626):1777–82.

Hotz, V. J., McElroy, S. W., and Sanders, S. G. (2005). Teenage childbearing and its life cycle consequences: Exploiting a natural experiment. *The Journal of Human Resources*, 40(3):683–715.

Hueston, W. J., McClaflin, R. R., and Claire, E. (1996). Variations in cesarean delivery for fetal distress. *The Journal of Family Practice*, 43(5):461–467.

Hyde, M. J. and Modi, N. (2012). The long-term effects of birth by caesarean section: The case for a randomised controlled trial. *Early Human Development*, 88(12):943–949.

Hyde, M. J., Mostyn, A., Modi, N., and Kemp, P. R. (2012). The health implications of birth by caesarean section. *Biological Reviews*, 87(1):229–243.

Jachetta, C. (2015). Cesarean sections and later health outcomes. http://www.christinejachetta.com/upload/jachetta_job_market_paper_2015.pdf. Accessed: 04-10-2016.

Jensen, V. M. and Wüst, M. (2015). Can caesarean section improve child and maternal health? the case of breech babies. *Journal of Health Economics*, 39:289 – 302.

Jeon, S.-H. and Pohl, R. V. (2017). Health and work in the family: Evidence from spouses' cancer diagnoses. *Journal of Health Economics*, 52:1 – 18.

Jiménez-Martín, S., Labeaga, J. M., and Martínez Granado, M. (1999). Health status and retirement decisions for older European couples. IRISS Working Paper Series 1999-01, IRISS at CEPS/INSTEAD.

Johnson, E. M. and Rehavi, M. M. (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, 8(1):115–41.

Jones, A. M., Rice, N., and Zantomio, F. (2019). Acute health shocks and labour market outcomes: Evidence from the post crash era. *Economics Human Biology*, page 100811.

Keag, O. E., Norman, J. E., and Stock, S. J. (2018). Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis. *PLOS Medicine*, 15(1):1–22.

Klasko, S. K., Cummings, R. V., Balducci, J., DeFulvio, J. D., and Reed, J. F. (1995). The impact of mandated in-hospital coverage on primary cesarean delivery rates in a large nonuniversity teaching hospital. *American Journal of Obstetrics and Gynecology*, 172(2):637 – 642.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2019a). Child Penalties Across Countries: Evidence and Explanations.

Kleven, H., Landais, C., and Søgaard, J. E. (2019b). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, 11(4):181–209.

Knip, M. and Simell, O. (2012). Enviornmental Triggers of Type 1 Diabetes. *Cold Spring Harbor perspectives in medicine*, 2(7):a007690.

Kuhle, S., Kirk, S. F. L., Ohinmaa, A., and Veugelers, P. J. (2011). Comparison of ICD code-based diagnosis of obesity with measured obesity in children and the implications for health care cost estimates. *BMC Medical Research Methodology*, 11:173.

Kvist, A. P., Nielsen, H. S., and Simonsen, M. (2013). The importance of children's adhd for parents' relationship stability and labor supply. *Social Science Medicine*, 88:30 – 38.

Lefèvre, M. (2014). Physician induced demand for c-sections: does the convenience incentive matter? Health, Econometrics and Data Group (HEDG) Working papers, HEDG, c/o Department of Economics, University of York.

Lenhart, O. (2019). The effects of health shocks on labor market outcomes: evidence from uk panel data. *The European Journal of Health Economics*, 20(1):83–98.

Leopold, T. (2018). Gender differences in the consequences of divorce: A study of multiple outcomes. *Demography*, 55(3):769–797. 29654601[pmid].

Leung, D. Y., Boguniewicz, M., Howell, M. D., Nomura, I., and Hamid, Q. A. (2004). New insights into atopic dermatitis. *The Journal of Clinical Investigation*, 113:651–657.

Lindeboom, M., Llena-Nozal, A., and van der Klaauw, B. (2016). Health shocks, disability and work. *Labour Economics*, 43:186 – 200. Health and the Labour Market.

Lobstein, T., Baur, L., and Uauy, R. (2004). Obesity in children and young people: a crisis in public health. *Obesity Reviews*, 5(s1):4–85.

Lundberg, S. and Rose, E. (2000). Parenthood and the earnings of married men and women. *Labour Economics*, 7(6):689 – 710.

Lundborg, P., Plug, E., and Rasmussen, A. W. (2017). Can Women Have Children and a Career? IV Evidence from IVF Treatments. *American Economic Review*, 107(6):1611–1637.

Malin, G. L., Morris, R. K., and Khan, K. S. (2010). Strength of association between umbilical cord pH and perinatal and long term outcomes: systematic review and meta-analysis. *BMJ: British Medical Journal*, 340:c1471.

Marcos, J. C. M. (2008). Cesarea a Demanda. Sociedad Española de Ginecología y Obstetricia.

McCloskey, L., Petitti, D. B., and Hobel, C. J. (1992). Variations in the use of cesarean delivery for dystocia: lessons about the source of care. *Med Care*, 30(2):126–135.

Meyer, B. D. and Mok, W. K. (2019). Disability, earnings, income and consumption. *Journal of Public Economics*, 171:51 – 69. Trans-Atlantic Public Economics Seminar 2016.

Miller, A. R. (2011). The effects of motherhood timing on career path. *Journal of Population Economics*, 24(3):1071–1100.

Ministerio de Sanidad, Servicios Sociales e Igualdad (2014). Costes hospitalarios - contabilidad analítica. Spanish Ministry of Health, Social Services and Equality. http://www.msssi.gob.es/estadEstudios/estadisticas/inforRecopila ciones/anaDesarrolloGDR.htm. Accessed: 13-01-2017.

Ministerio de Sanidad, Servicios Sociales e Igualdad (2015). Informe anual del sistema nacional de salud. Spanish Ministry of Health, Social Services and Equality. http://www.msssi.gob.es/estadEstudios/estadisticas/sisInfSanSNS/ tablasEstadisticas/ InfAnSNS.htm. Accessed: 13-01-2017.

Ministerio de Sanidad y Política Social (2009). Informe maternidad hospitalaria: Estándares y recomendaciones. Spanish Ministry of Health and Social Policy. http://www.msc.es/organizacion/sns/planCalidadSNS/docs/ AHP.pdf. Accessed: 14-12-2017.

Ministerio de Trabajo y Asuntos Sociales (1997). BOE-A-1997-26857. https://www.boe.es/diario_boe/txt.php?id=BOE-A-1997-26857.

Ministry of Social Affairs and Health (2013). Health Care in Finland. Ministry of Social Affairs and Health.

Mishanina, E., Rogozinska, E., Thatthi, T., Uddin-Khan, R., Khan, K. S., and Meads, C. (2014). Use of labour induction and risk of cesarean delivery: a systematic review and meta-analysis. *Canadian Medical Association Journal*, 186(9):665–673.

Mühlrad, H. (2017). Cesarean section for high-risk births: short and long term consequences for breech births. https://www.dropbox.com/s/xto1qlsf1qtxf7d/C-section_Muhlrad.pdf. Accessed: 26-12-2017.

Neu, J. and Rushing, J. (2011). Cesarean versus vaginal delivery: Long-term infant outcomes and the hygiene hypothesis. *Clinics in Perinatology*, 38(2):321 – 331. Delivery After Previous Cesarean.

NICE (2008). Inducing labour. NICE guideline CG70, National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/cg70. Accessed: 08-06-2017.

NICE (2011). Cesarean section. NICE clinical guideline CG190, National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/cg132/chapter/1-guidance. Accessed: 06-12-2017.

NICE (2014). Intrapartum care for healthy women and babies. NICE guideline CG190, National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/cg190. Accessed: 06-02-2017.

NICE (2016). Deciding whether to offer caesarean section. NICE pathways, National Institute for Health and Care Excellence. http://pathways.nice.org.uk/pathways/caesarean-section. Accessed: 04-10-2016.

Nielsen, H. S., Smith, N., and Çelikaksoy, A. (2009). The effect of marriage on education of immigrants: Evidence from a policy reform restricting marriage migration. *Scandinavian Journal of Economics*, 111(3):457–486.

Nix, E. and Andresen, M. E. (2019). What Causes the Child Penalty? Evidence from Same Sex Couples and Policy Reforms. Discussion Papers 902, Statistics Norway, Research Department.

OECD (2013). *Health at a Glance 2013: OECD Indicators*. OECD Publishing, Paris.

OECD (2017). *Health at a Glance 2017: OECD Indicators*. OECD Publishing, Paris.

Oreopoulos, P., Stabile, M., Walld, R., and Roos, L. L. (2008). Short-, medium-, and long-term consequences of poor infant health: An analysis using siblings and twins. *Journal of Human Resources*, 43:88–138.

Otamiri, G., Berg, G., Ledin, T., Leijon, I., and Lagercrantz, H. (1991). Delayed neurological adaptation in infants delivered by elective cesarean section and the relation to catecholamine levels. *Early Human Development*, 26(1):51 – 60.

Ottosson, F., Brunkwall, L., Ericson, U., Nilsson, P. M., Almgren, P., Fernandez, C., Melander, O., and Orho-Melander, M. (2018). Connection between bmi-related plasma metabolite profile and gut microbiota. *The Journal of Clinical Endocrinology & Metabolism*, 103(4):1491–1501.

Page, M. E. and Stevens, A. H. (2004). The economic consequences of absent parents. *The Journal of Human Resources*, 39(1):80–107.

Paull, G. (2008). Children and women's hours of work*. *The Economic Journal*, 118(526):F8–F27.

Pertold-Gebicka, B., Pertold, F., and Datta Gupta, N. (2016). Employment adjustments around childbirth.

Peters, L. L., Thornton, C., de Jonge, A., Khashan, A., Tracy, M., Downe, S., Feijen-de Jong, E. I., and Dahlen, H. G. (2018). The effect of medical and operative birth interventions on child health outcomes in the first 28 days and up to 5 years of age: A linked data population-based cohort study. *Birth*, pages 1–11.

Powers, E. T. (2003). Children's Health and Maternal Work Activity: Estimates under Alternative Disability Definitions . *Journal of Human Resources*, XXXVIII(3):522–556.

Prior, E., Santhakumaran, S., Gale, C., Philipps, L. H., Modi, N., and Hyde, M. J. (2012). Breastfeeding after cesarean delivery: a systematic review and meta-analysis of world literature. *The American Journal of Clinical Nutrition*, 95(5):1113–1135.

Reed, S. D., Lee, T. A., and McCrory, D. C. (2004). The Economic Burden of Allergic Rhinitis. *PharmacoEconomics*, 22(6):345–361.

Rellstab, S., Bakx, P., García-Gómez, P., and van Doorslaer, E. (2019). The kids are alright - labour market effects of unexpected parental hospitalisations in the netherlands. *Journal of Health Economics*, page 102275.

Saccone, G. and Berghella, V. (2015). Induction of labor at full term in uncomplicated singleton gestations: a systematic review and meta-analysis of randomized controlled trials. *American Journal of Obstetrics and Gynecology*, 213(5):629 – 636.

Scudellari, M. (2017). News feature: Cleaning up the hygiene hypothesis. *Proceedings of the National Academy of Sciences*, 114(7):1433–1436.

SEGO (2005). Monitorización fetal intraparto. *Progresos de Obstetricia y Ginecología*, 48(4):207–216.

Sevelsted, A., Stokholm, J., Bønnelykke, K., and Bisgaard, H. (2015). Cesarean section and chronic immune disorders. *Pediatrics*, 135(1):e92–e98.

Shurtz, I. (2013). The impact of medical errors on physician behavior: Evidence from malpractice litigation. *Journal of Health Economics*, 32(2):331 – 340.

Sieppi, A. and Pehkonen, J. (2019). Parenthood and gender inequality: Population-based evidence on the child penalty in finland. *Economics Letters*, 182:5 – 9.

Sigle-Rushton, W. and Waldfogel, J. (2007). Motherhood and women's earnings in anglo-american, continental european, and nordic countries. *Feminist Economics*, 13(2):55–91.

Skiold, B., Petersson, G., Ahlberg, M., Stephansson, O., and Johansson, S. (2017). Population-based reference curve for umbilical cord arterial pH in infants born at 28 to 42 weeks. *Journal of Perinatology*, 37(3):254–259.

Skoner, D. P. (2001). Allergic rhinitis: Definition, epidemiology, pathophysiology, detection, and diagnosis. *Journal of Allergy and Clinical Immunology*, 108(1, Supplement):S2 – S8.

Snaebjorn, G. and Steingrimsdottir, H. (2019). The Long-Term Impact of Children's Disabilities on Families.

Spetz, J., Smith, M. W., and Ennis, S. F. (2001). Physician incentives and the timing of cesarean sections: evidence from California. *Medical care*, 39(6):536–550.

Spong, C. Y., Berghella, V., Wenstrom, K. D., Mercer, B. M., and Saade, G. R. (2012). Preventing the first cesarean delivery: Summary of a joint Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-fetal Medicine, and American College of Obstetricians and Gynecologists workshop. *Obstetrics and Gynecology*, 120(5).

Stabile, M. E. and Allin, S. (2012). The economic costs of childhood disability. *The Future of children*, 22 1:65–96.

Stock, J. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Andrews, D. W., editor, *Identification and Inference for Econometric Models*, pages 80–108. Cambridge University Press, New York.

Strachan, D. P. (1989). Hay fever, hygiene, and household size. *BMJ (Clinical research ed.)*, 299(6710):1259–60.

Sund, R. (2012). Quality of the Finnish Hospital Discharge Register: A systematic review. *Scandinavian Journal of Public Health*, 40(6):505–515.

Tao, B., Pietropaolo, M., Atkinson, M., Schatz, D., and Taylor, D. (2010). Estimating the cost of type 1 diabetes in the U.S.: A propensity score matching method. *PLoS ONE*, 5(7).

Thavagnanam, S., Fleming, J., Bromley, A., Shields, M. D., and Cardwell, C. R. (2008). A meta-analysis of the association between Caesarean section and childhood asthma. *Clinical & Experimental Allergy*, 38(4):629–633.

Tonei, V. (2019). Mother's mental health after childbirth: Does the delivery method matter? *Journal of Health Economics*, 63:182–196.

Torpy, J. M. (2010). Chronic diseases of children. *JAMA - Journal of the American Medical Association*, 303(7):682.

Trevisan, E. and Zantomio, F. (2016). The impact of acute health shocks on the labour supply of older workers: Evidence from sixteen european countries. *Labour Economics*, 43:171 – 185. Health and the Labour Market.

Tuomisto, L., Erhola, M., Luukkaala, T., Puolijoki, H., Nieminen, M., and Kaila, M. (2010). Asthma programme in Finland: Did the use of secondary care resources become more rational? *Respiratory Medicine*, 104(7):957 – 965.

Tynkkynen, L.-K., Chydenius, M., Saloranta, A., and Keskimäki, I. (2016). Expanding choice of primary care in finland: much debate but little change so far. *Health Policy*, 120(3):227 – 234.

van den Berg, G. J., Lundborg, P., and Vikström, J. (2017). The economics of grief. *The Economic Journal*, 127(604):1794–1832.

Wagstaff, A. (2007). The economic consequences of health shocks: Evidence from vietnam. *Journal of Health Economics*, 26(1):82 – 100.

Waldfogel, J. (1998). Understanding the ”family gap” in pay for women with children. *The Journal of Economic Perspectives*, 12(1):137–156.

Wasi, N., van den Berg, B., and Buchmueller, T. C. (2012). Heterogeneous effects of child disability on maternal labor supply: Evidence from the 2000 us census. *Labour Economics*, 19(1):139 – 154.

Webb, D. A. (2002). Time of day variation in rates of obstetric intervention to assist in vaginal delivery. *Journal of Epidemiology & Community Health*, 56(8):577–578.

WHO (2015). *WHO Statement on Caesarean Section Rates*. World Health Organization, Geneva.

Wier, L., Pfuntner, A., Maeda, J., Stranges, E., Ryan, K., Jagadish, P., Collins Sharp, B., and Elixhauser, A. (2011). Hcup facts and figures: statistics on hospital-based care in the united states, 2009. *Rockville, MD: Agency for Healthcare Research and Quality*.

Wolfe, B. L. and Hill, S. C. (1995). The effect of health on the work effort of single mothers. *The Journal of Human Resources*, 30(1):42–62.

Wright, S. M. and Aronne, L. J. (2012). Causes of obesity. *Abdominal Radiology*, 37(5):730–732.

Zanardo, V., Simbi, A., Franzoi, M., Soldá, G., Salvadori, A., and Trevisanuto, D. (2004). Neonatal respiratory morbidity risk and mode of delivery at term: influence of timing of elective caesarean delivery. *Acta Paediatrica*, 93(5):643–647.