The Long-Run Effects of Cesarean Sections

Ana María Costa-Ramón †
Universitat Pompeu Fabra & CRES

Mika Kortelainen
VATT Institute for Economic Research

Ana Rodríguez-González
Universitat Pompeu Fabra & CRES

Lauri Sääksvuori
National Institute for Health and Welfare (THL)

Preliminary draft, do not circulate

Abstract

Cesarean section is the most common surgical procedure in many countries around the world. Cesarean delivery for low-risk pregnancies is associated with several adverse health outcomes for infants and mothers. The interpretation of these correlations is, however, confounded due to the selection of birth mode. We use high quality administrative data which includes detailed birth and health records for all children born in Finland from 1990 to 2014 to study the causal effects of cesarean delivery on infants' long-term health. We show that physicians are more likely to perform C-sections during their regular shift on Fridays and working days that precede public holidays and use this variation as an instrument for unplanned C-sections. We supplement our instrumental variables estimates using variation within sibling pairs and across families where the second child is born either by unplanned C-section or vaginal delivery. Our results suggest that avoidable unplanned C-sections increase the risk of asthma, but do not affect the probability of being diagnosed with other immune and metabolic disorders previously associated with C-sections.

Keywords: Cesarean section, infant health, time variation, instrumental variables.

JEL Codes: I10.

†Corresponding author: Ana María Costa-Ramón. Department of Economics, Universitat Pompeu Fabra. C/ Ramon Trias Fargas 25-27, 08005 Barcelona (Spain). Email: anamaria.costa@upf.edu

1 Introduction

There is little doubt that prenatal health and early childhood circumstances can have life-long consequences on mortality, morbidity, and human capital development. The theory of fetal origins of adult disease has proven to describe a surprisingly general phenomenon. The long-term effects of prenatal health 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 neonates. There is accumulating evidence that many common interventions around birth are associated with children's long-term physical health and human skill formation (Peters et al., 2018). However, the causal nature of these relationships has received little attention so far. Most notably, the literature associates births by cesarean sections to a wide variety of worse short- and long-term health outcomes.

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 to adjust the human immune system to react appropriately 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, differential exposure to bacteria of babies born by cesarean section would affect the development of their immune system and make them more prone to immune-mediated diseases, such as asthma, allergies, type 1 diabetes, and other metabolic disorders.

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 design. Crucially, such study must not be limited to short-term health indicators but must cover also long-term outcomes, so that it is possible to detect effects of C-sections that are mediated by alterations of the immune system. The rapidly growing incidence of cesarean sections across the globe[1] suggests that even small increases in mortality and

[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 section are reported in many of the world's most

morbidity due to C-sections would lead to large reductions in life expectancy and substantial welfare effects.

This paper provides new evidence on the effect of avoidable cesarean sections on several policy relevant health outcomes using large and precise administrative data registers. 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 that affects the rate of unplanned cesarean sections. We find that the probability of unscheduled C-section increases substantially during the normal working hours (8am – 4pm) on working days that precede a leisure day – i.e, Fridays and holiday eves – while mothers giving birth at these times are similar in a large set of observable characteristics.

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 a greater use of more discretionary diagnoses, while it is unrelated to medical emergencies. Moreover, we observe that physician demand for leisure does not affect groups of mothers (medical professionals) whose mode of delivery is shown by the literature not to respond to doctors' incentives. Taken together, our data lend strong support for the contention that the excess numbers of cesarean deliveries observed during the normal working hours on days that precede a leisure day are largely due to physicians' incentives. This enables us to use the observed time-variation as an instrument for C-section.

We investigate the effects of cesarean sections on infant outcomes using a rich 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 child health. We focus on the 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). Understanding and quantifying the potential contribution of C-sections to the development of these diseases is important, given their high direct and indirect

populous countries including among others China (41.3 percent in 2016) and Brazil (55.6 percent in 2015). For a recent review of disparities in C-section use around the world see Boerma et al. (2018).

costs.[2]

To the best of our knowledge, this is the first paper to study the causal impact of avoidable cesarean sections on later infant health.[3] There is a nascent literature in economics that studies the causal relationship between cesarean sections and various short-term infant outcomes. Jachetta (2015) explores the effect of cesarean delivery on hospitalizations using variation in medical malpractice insurance premia by Metropolitan Statistical Area as an instrument for the C-section rate. Card et al. (2018) study the health effects of cesarean deliveries within one year after birth using relative distance from a mother's home to hospitals with high and low rates of cesarean sections as an instrument for mode of delivery. Both studies find an increased rate of hospital visits for cesarean born babies, mostly due to respiratory conditions.

In this paper, we advance the existing knowledge by studying detailed diagnoses in the longer term, focusing on unscheduled c-sections. Moreover, our identification allows us exploit variation in avoidable interventions which is unrelated to maternal characteristics, and to compare mothers giving birth in the same hospital, thus abstracting from other factors that could change across hospitals or geographical areas.

In a recent study, Costa-Ramón et al. (2018) investigate the effects of cesarean sections on neonatal health using time variation in unplanned cesarean sections rates during 24-hour duties of obstetricians in Spanish hospitals. Their results show that C-sections have a significant negative impact on Apgar scores. However, this effect is relatively small and does not translate into more severe outcomes. In line with previous evidence, we find that C-sections have a negative impact on Apgar scores but do not lead to an increased mortality risk. We also find a significant increase in the probability of intensive care unit admission and assisted ventilation.

[2]The total cost of asthma in the working age population was estimated to be US $24.7 billion during 1999-2002 in Europe (Global Asthma Network, 2018). The two other atopic diseases we investigate also 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 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 only in lifetime medical costs in the US (Finkelstein et al., 2014).

[3]The work by Jensen and Wüst (2015) and Mühlrad (2017) examine the short- and long-term impacts of medically necessary C-sections for the particular case of breech babies. They find largely positive health effects for this high-risk group. However, the effects of medically necessary C-sections might be very different from the effects of unnecessary C-sections that we investigate in this paper.

Importantly, our long-run results contribute to obtaining a more complete picture of the effect of C-sections, given that alterations of the immune system need not be visible at birth.

Our instrumental variable estimates suggest that avoidable C-sections increase the probability of an asthma diagnosis from early childhood onwards. This effect is of clinical and economic relevance, and consistent with the hypothesis that mode of delivery can impact the development of the immune system. However, we do not find consistent evidence that cesarean sections affect the probability of other atopic diseases, type 1 diabetes, or obesity diagnoses.

We complement these results using variation within and between sibling pairs: we compare the health gap between the second and the first child in families where the second child is born by unplanned C-section with respect to families where both children were born by vaginal delivery. This empirical strategy controls for all time-invariant unobserved heterogeneity across families. We also control for a large set of maternal, pregnancy, and delivery characteristics that could differ between siblings, even though we cannot rule out all potential time-varying confounders. We argue, however, that these confounders, if any, would make our estimates negatively biased, since the second siblings born by C-sections are likely to be negatively selected with respect to second siblings born by vaginal delivery. Therefore, if within-family results show that C-sections do not have an impact on a given outcome, this would be strong suggestive evidence for the lack of an effect.

The results from our supplementary empirical strategy support our findings. Unplanned C-sections increase the risk of childhood asthma, but we find evidence of no effect on other atopic diseases, type 1 diabetes, or obesity. We provide several sensitivity checks that suggest that the effect on asthma is unlikely to be explained by selection alone. Overall, our results paint a more nuanced picture about the long-term consequences of cesarean deliveries than existing evidence based mostly on associations. Our results suggest that C-sections cause a much narrower spectrum of diseases than currently hypothesized.

The paper is structured as follows. Section 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 3 introduces the data, provides key descriptive statistics and lays out our econometric approach. Section 4 reports our main results. Section 5 presents robustness checks and additional evidence to support our main conclusions. The last section concludes.

2 Background

2.1 Mechanisms

A large body of scientific literature documents the developmental origins of health and disease. The process of labor can be seen as one crucial step in the initiation of adaptation to 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). Although 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 long-term effects on health and anomalies in human development. In the following, we summarize some of the most widely acknowledged findings from epidemiological studies 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 lack of early childhood exposure to infection 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, dubbed as the old friends hypothesis, have challenged the role of infectious pathogens and highlight the importance of an 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 of newborns 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). Given the negative association between cesarean sections and breastfeeding initiation, this might be another channel explaining the differences in microbiota by type of birth (Prior et al., 2012; Hyde et al., 2012).

In addition to the effect on microbiota and gut colonization, there are other short-term physiological differences between cesarean and vaginally delivered infants. These differences include impaired lung function and altered behavioral responses to stress. The most common cause of respiratory distress among newborns is transient tachypnea due to the presence of retained lung fluid. In vaginal births, the process of labor helps to expel the amniotic fluids from the lungs. However, this process is compromised when the baby is born by a C-section, and the presence of fluid in the lungs after birth is more common. This increased respiratory morbidity at birth is hypothesized to increase long-term morbidity (Hyde et al., 2012). Finally, the process of labor is associated with the release of catecholamines which cause physiological changes that facilitate fetal adaptation to extrauterine stress. Research suggests that cesarean delivery is associated with lower levels of catecholamines than vaginal delivery (Hyde et al., 2012). This lack of catecholamine may, in turn, cause further alterations in the programming of the central nervous system with corresponding long-term effects (Otamiri et al., 1991).

The potential mechanisms previously described are consistent with the reported associations between cesarean delivery and adverse infant outcomes in a large number of observational studies. These studies relate cesarean deliveries to a marked increase in the susceptibility of multiple immune and metabolic conditions. Although associations have been explored with a broad array of such diseases, recent meta-analysis conclude that C-sections are most robustly related to asthma and other atopic diseases, type 1 diabetes and obesity (Blustein and Liu, 2015; Keag et al., 2018; Cardwell et al., 2008; Thavagnanam et al., 2008; Li et al., 2013; Peters et al., 2018; Bager et al., 2008; Magne et al., 2017).[4] Despite the abundant number of papers reporting significant associations and plausible biological mechanisms, the causal nature and clinical relevance of these relationships remains largely unknown.[5]

2.2 Classification of Cesarean Sections

Cesarean section is worldwide one of the most commonly performed major surgeries (Boerma et al., 2018). Cesarean sections are performed for several indications at different stages of the pregnancy. Conventionally, 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 overwhelming 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.

The large majority of 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 hospital. Unscheduled C-sections can be further 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 more discretion to the clinician. Slow progression of labor or cephalopelvic

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

[5]Hyde et al. (2012) summarize evidence from 14 RCTs that compare the effects of cesarean and vaginal deliveries on infant health. However, all of 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. In the same vein, 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 more 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.

disproportion are examples of diagnoses that may require an unplanned non-urgent cesarean section, and where wide variation is found among clinicians (Barber et al., 2011; Fraser et al., 1987). Given that, conditional on maternal characteristics, emergencies should be uniformly distributed during the day, we expect any observed time variation in unplanned C-sections to be driven by these non-emergency cesareans. Moreover, for a subsample of births our data contains the specific indication registered by the medical team to justify the C-section. This allows us to verify that the peaks in unplanned C-sections are coupled with the use of more discretionary diagnoses.

2.3 Institutional Context

Finland has universal public health coverage, and comprehensive pre- and postnatal care services are included in the publicly provided services. There are no private medical institutions running maternity wards. Consequently, practically 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 any pre-assigned midwife or physician for the delivery. Midwives take care of the delivery in all hospitals, but physicians have the ultimate responsibility for obstetric care and there are no midwife-led delivery units. Physicians decide on the type of delivery and perform C-sections. The C-section rate is low from an international perspective: 15.5% in 2015 (OECD, 2017).

The regular working shift for a physician is from 8 am to 4 pm from Monday to Friday. When on duty, physicians must work for 24 hours (from 8 am to 8 am), except on weekends when physicians must stay in the hospital for 25 hours (from 8 am to 9 am on next day).[6] Midwives follow the same rotation regardless of the type of day and work in three shifts of around 8

[6]Even though the system has changed in recent years, during most years covered in our data, small hospitals with less than 1000 annual births decided their on call arrangements autonomously. 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.

hours.[7]

3 Data and methods

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, and 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). The 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:[8] our resulting sample consists of 392,560 deliveries that took place from 1990 to 2014. For the within-family 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 3.2.2): the resulting sample consists of 645,292 children from 322,646 sibling pairs. There are 43 hospitals in our sample. Table A1 shows summary statistics for all births in Finland in this period.

We match the Finnish Medical Birth Register to the Finnish Hospital Discharge Register, which contains information about the diagnosed medical conditions, medical operations, and the date of diagnoses. This hospital register contains all inpatient consultations in Finland from 1988 to

[7] 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.

[8] We follow the literature by focusing on first births, which also allows us to keep just one birth per mother, thus abstracting from another 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.

2013. All diagnoses are coded using the International Classification of Diseases (ICD) tool.[9] From 1998, the data also include all outpatient visits to hospitals.[10]

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 (in particular, an indicator for Apgar scores below 7)[11], admission to intensive care unit (ICU), need of assisted ventilation, and early neonatal mortality (defined as neonatal death in the first week of life). Second, we study longer term outcomes using detailed inpatient and outpatient diagnosis data from the Finnish Hospital Discharge Register. We use primary diagnoses.[12] 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 A2 in the appendix provides more detail about each of these diagnoses.

3.2 Empirical strategy

We estimate the impact of a cesarean delivery on child's health, both at birth and at 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. We would thus like to estimate the following equation:

$$Y_i = \beta_0 + \beta_1 CS_i + \beta_2 X_i + \delta_m + \lambda_y + \phi_h + \epsilon_i \tag{1}$$

[9] Diagnoses for years from 1987 to 1995 are recorded using ICD-9 classification. Diagnoses from 1996 onwards are recorded using ICD-10 classification.

[10] 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 evaluate more specifically to what extent our data is able to identify the individuals with a certain diagnosis in the Results section.

[11] 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.

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

where X_i is a vector of covariates that includes information on the gender of the baby, the mother's marital status, nationality, socioeconomic status, age, smoking status, pregnancy characteristics;[13] and δ_m, λ_y, and ϕ_h are fixed effects for the month, year, and hospital of birth, respectively.

However, the estimation of equation (1) is likely to provide biased estimates of β_1 due to potential selection by birth mode. Figure A2 in the appendix shows that mothers getting a C-section and their infants are very different from those who undergo a vaginal delivery. To study the causal effects of cesarean delivery on health, we exploit two different identification strategies: an instrumental variable and a sibling fixed effect model. In what follows, we explain each method in greater detail.

3.2.1 IV approach: variation by type of day

Our first approach will be to exploit the observed higher likelihood of giving birth by C-section during the normal working shift on pre-leisure days compared to regular working days. We will use the interaction of the type of day and shift of birth as an instrument for the mode of delivery.

Figure 1 presents the predicted probability of having an unplanned C-section delivery by hour and type of day. We adjust for hospital, month, and year of birth fixed effects. Figure 1a plots the distribution of the C-section rate over a 24-hour cycle for working days that precede a leisure day – that is, Fridays or other working days preceding a public holiday – compared to other working days.[14] We find that substantially more C-sections are performed during regular working hours in days that precede a leisure day compared to the rest of working days. Figure 1b presents the predicted probability of having an unplanned C-section by work shift and type of day. It clearly shows that the gap in C-section rates between pre-leisure days and other working days emerges only during the regular working shift (from 8 am to 4 pm).

[13] We include 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, gestational weeks, induced labor, prostaglandin pre-induction, epidural use, and laughing gas anesthesia.

[14] See table A3 for an example of public holidays in Finland. A given Friday is not considered a pre-leisure day if it is a holiday itself.

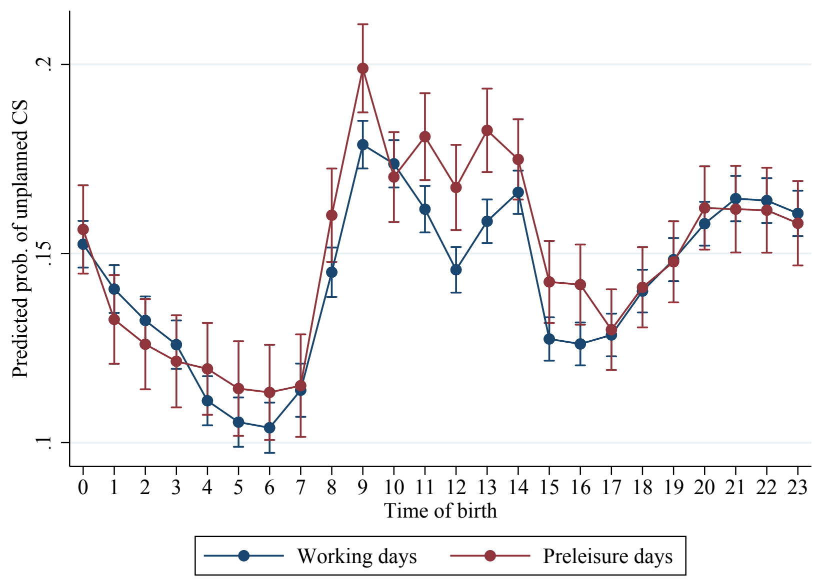
Importantly, we find that the excess C-sections performed in days that precede a leisure day are not exclusively driven by advancing births that would have been by cesarean in any event. If this was the case, we would observe that this relative surge in C-sections is matched by a relative drop compared to other working days later on, either on evening hours of the pre-leisure day of during the next day (the leisure day). However, we do not observe this relative fall, either during the evening (Figure 1a) or during the leisure day (see Figure A1 in the Appendix[15]). These observations imply 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 in pre-leisure days. Halla et al. (2016) exploit this variation in an instrumental variable framework to study the impact of mode of delivery on maternal subsequent 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 and manifests as a decreasing willingness to wait for the normal progression of vaginal deliveries. A cesarean section takes on average 30-75 minutes, and is perceived as a relatively easy intervention with low complication rates (NICE, 2011), while the average duration of labor for first-time mothers who have a vaginal birth is 11 hours (NICE, 2014).

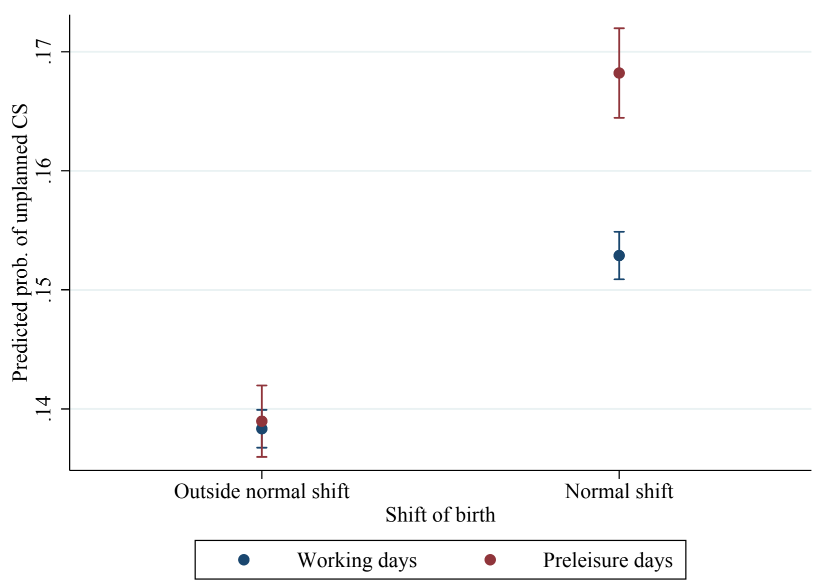
We provide two pieces of complementary evidence to validate that the excess rate of C-sections is not driven by medical factors. First, we build on previous evidence that some medical diagnoses linked to a cesarean birth are more discretionary than others. Dystocia (prolonged or obstructed labor), one of the most common indications for primary cesarean section, is especially 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 A4) suggest that giving birth during the regular hour on pre-leisure day increases the

[15] 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 1: Predicted probability of unplanned C-section by time and type of day



(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 includes working days that precede a Finnish public holiday or a weekend, while working days includes the rest of working days. Sample is restricted to singleton first births which are either unscheduled C-sections or vaginal births.

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.[16]

Our second piece of evidence builds on the literature showing that physician mothers are less likely to get 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 A5) confirm 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. In contrast, this effect is found for a group of non-medical mothers with a similar level of education[17]

We thus use this variation and adopt an instrumental variable approach, with the following first stage:

$$CS_i = \gamma_0 + \gamma_1 NS_i + \gamma_2 Preleisure_i + \gamma_3 NS_i \times Preleisure_i + \gamma_4 X_i + \delta_m + \lambda_y + \phi_h + v_i \qquad (2)$$

and the corresponding second stage:

$$Y_i = \alpha_0 + \alpha_1 NS_i + \alpha_2 Preleisure_i + \alpha_3 \widehat{CS}_i + \alpha_4 X_i + \delta_m + \lambda_y + \phi_h + \varepsilon_i \qquad (3)$$

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 (3) are the

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

[17]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.

predicted C-sections from the first stage, X_i is the vector of individual controls,[18] and δ_m, λ_y and ϕ_h are month, year, and hospital of birth fixed effects, respectively. Therefore, the interaction between regular working hours and a day preceding a leisure day will serve as an instrument. We expect a positive $\widehat{\gamma}_3$ due to increasing physician demand for leisure on days preceding a weekend or public holidays. 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).

Given the binary nature of most of the health outcomes and the endogenous variable, we also estimate bivariate probit models. According to Chiburis et al. (2012), linear IV estimates are particularly uninformative when treatment probabilities are low and the model includes additional covariates, conditions that apply in our context. Angrist and Pischke (2009) consider bivariate probit models to be a valid and "harmless" approach. Thus, we will also report the marginal effects of this model for our variables of interest.[19]

For this to be a valid IV strategy, three conditions must be met. 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 in 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).

Regarding the first condition, Table 1 shows the results from the estimation of the first stage: in column (1) just with month, year, and hospital fixed effects, and in column (2) with a richer set of controls. The estimates show that being born during the normal shift increases the probability of the delivery being a C-section for all working days, and being born during this time on pre-leisure days increases the probability of C-section by 1.5 percentage points. The first stage

[18]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.

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

F-statistics are larger than 25 in both 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 the instrument is weak.

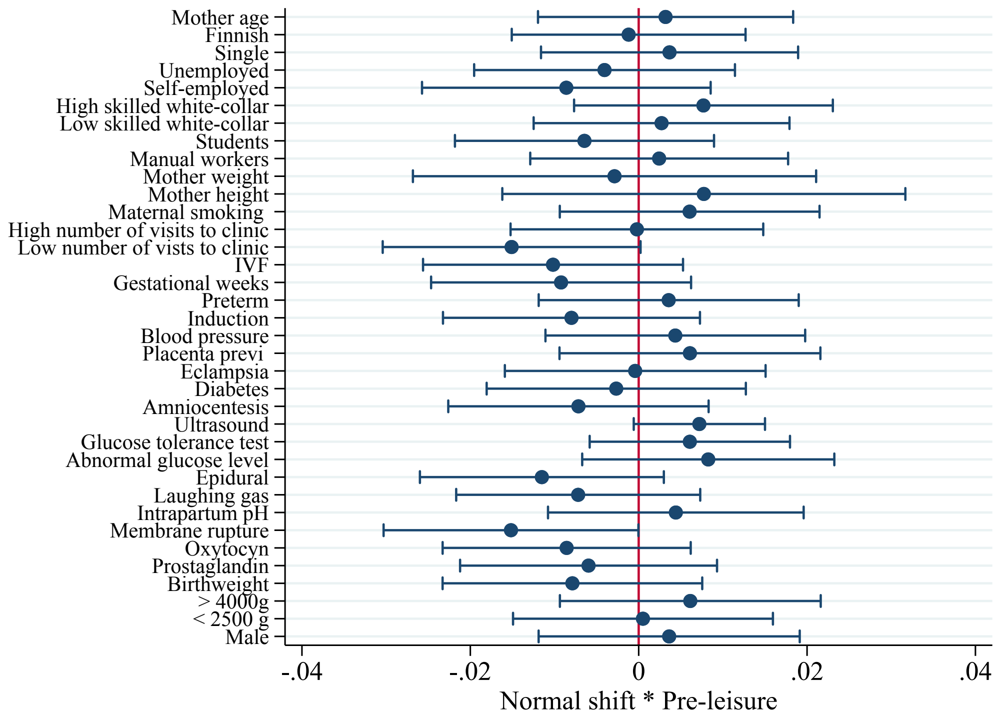
Table 1: First stage

	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

Notes: This table shows estimates from the first stage (see equation (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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$

While the second assumption cannot be empirically tested, Figure 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. Finally, with regard to the exclusion restriction, since we focus our comparison on births that take place on working days, we expect hospital resources and quality of care to be constant, and we would thus not expect anything else to change that could affect the health of the child. Moreover, for it to compromise our identification strategy, any such change would

Figure 2: Relation of the instrument with 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 in working days.

need to happen on pre-leisure days but only during the regular shift. In section 5.1 we provide supplementary analyses that further reinforce the credibility of this assumption.

Our IV will allow us to identify the local average treatment effect (LATE), that is, the effect of C-sections for infants whose mothers' mode of delivery is sensitive to the subjective assessment of the physician. More specifically, we capture cases where the type of day affects the decision of the doctor to perform a C-section during the normal shift. The counterfactual for these cases is unlikely to be exclusively a later cesarean section, given that we do not find a relative drop in C-sections later on pre-leisure days (outside the normal shift) or during the following day, as previously discussed. This LATE will not be informative of the effect of medically indicated C-sections, given that medical necessity should not be affected by leisure incentives, and will probably not capture the effect of unplanned C-sections for babies who had a very fast delivery, leaving no room for physician discretion.

3.2.2 Within-family analysis

Our second empirical strategy consists of applying a difference-in-differences approach to a sample of siblings. We restrict the sample to families where the older sibling was born by vaginal delivery, and compare the health gap between the first and the second born child in families where the latter was born by an unplanned C-section with respect to families whose second-child was born by vaginal delivery. This allow us to control for all time-invariant unobserved heterogeneity at the family level and to control for the effect of birth order. Numerous papers have used siblings fixed-effects models 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; Black et al., 2016; Aizer et al., 2016).

We will estimate the following equation:

$$Y_{if} = \alpha_0 + \alpha_1 Secondborn_{if} + \alpha_2 Secondborn_{if} \times CS_{if} + \alpha_3 X_{if} + \gamma_f + +\delta_m + \lambda_y + \phi_h + \varepsilon_{if} \quad (4)$$

where Y_{if} is the health outcome of child i of family f, $Secondborn_{if}$ is a dummy 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 (3), except for maternal characteristics that are time-invariant, and diagnoses during delivery (prolonged and obstructed labor[20]); and γ_f, δ_m, λ_y and ϕ_h are family, month, year, and hospital of birth fixed effects, respectively.[21] We cluster standard errors at the family level. Our parameter of interest is α_2, which identifies 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.

[20] 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.

[21] 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.

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 would be a very selected sample, given the very high probability of having a repeated C-section.[22] 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.

Although we are able to control for a large set of observable characteristics that could change between both pregnancies, some time-varying unobservable differences could still remain. If this is the case, we would expect siblings born by C-section to be negatively selected: compared to their vaginally-delivered older siblings, second children born by C-section might have had some problems either during pregnancy or during the delivery that we cannot observe in our data and that lead to the C-section. This would cause our estimates to be negatively biased. Having the direction of the bias in mind, if we find that siblings born by C-section do not have a higher probability of a given diagnosis, this will be reassuring evidence of no effect of C-sections on that disease. In section 5.2 we assess in more detail the extent to which our results can be explained by selection.

4 Results

4.1 Neonatal outcomes

We first estimate the impact of C-sections on neonatal outcomes. Table 2 shows estimates from four different methods: OLS results in the first panel; 2SLS estimates in the second; bivariate probit marginal effects in the third, and coefficients from the siblings fixed effects model in the last one.

OLS results confirm previous findings in the literature, with cesarean sections being associated with worse outcomes at birth and with higher neonatal mortality.[23] The 2SLS estimates are

[22]Only after 2010 did The American College of Obstetricians and Gynaecologists (ACOG) encourage 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).

[23]The OLS estimation is ran in a sample that only excludes planned C-sections and births for which we do not

not significant for any of the indicators. However, we cannot reject that there is a (potentially large) effect, given the magnitude of the coefficient and the large standard errors. As discussed by Chiburis et al. (2012) and Andrabi et al. (2013) confidence intervals obtained from linear IV estimates are particularly large when treatment probabilities are low and the model includes additional covariates.

The bivariate probit coefficients are much more precisely estimated, but consistent with the 2SLS results, given that they are included in the confidence intervals of the point estimates. The bivariate probit results suggest that unplanned C-sections increase the probability of the newborn having a low Apgar score (Apgar lower than 7), being admitted to the intensive care unit, and receiving assisted ventilation. The magnitude of the coefficients is similar to that of OLS estimates. However, we do not find a significantly increased mortality risk at seven days after birth. The results from the family fixed effects models corroborate this finding, with similarly-sized coefficients. All in all, our results suggest that unplanned C-sections have a negative impact on neonatal health, but that this effect does not translate into a higher probability of early neonatal mortality.

4.2 Later infant health

Table 3 shows OLS (first panel), two-stage least squares coefficients (second panel), bivariate probit marginal effects (third panel) and estimates from siblings fixed-effects model (fourth panel), for the coefficient of unplanned C-sections on the probability of the infant having a given diagnosis by age, for ages 5 and 10. In particular, we analyze health conditions that have been extensively documented in the medical literature as being positively associated with cesarean deliveries: type 1 diabetes, obesity, asthma, and other atopic diseases (atopic dermatitis and allergic rhinitis). It is important to note that, given that we study health outcomes of infants born from 1990 to 2014, our sample decreases as we consider older ages. More detailed estimates year by year up to age 15 are shown in Figures 3 and 4.

observe parity. The specification includes the full set of controls and fixed effects described in equation (1), as well as controls for birth order.

Table 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
Siblings f.e.	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

Notes: This table shows the estimates of the effect of an unplanned CS on different neonatal health indicators by OLS, 2SLS, bivariate probit and siblings fixed effects estimation (see equations (2), (3), and (4)). Specifications as detailed in section 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 siblings F.E. panel. First-stage F statistic from 2SLS and bivariate probit specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$

The OLS estimates suggest that cesarean sections are associated with a higher probability of asthma, obesity, and atopic diseases, but in our data we do not detect that C-section born babies have a higher probability of a type 1 diabetes diagnosis. These findings are again in line with previous studies documenting negative associations between cesarean sections and metabolic and immune-related conditions (Blustein and Liu, 2015; Magne et al., 2017).

The 2SLS results for type 1 diabetes suggest that unplanned C-sections significantly increase the probability of this diagnosis before age 5, although the effect is not significant by age 10. The magnitude of the estimate is quite large, but very imprecise: it suggests an increase in the probability of type 1 diabetes of 9 percentage points, but is consistent with an increase ranging from 6.3 to 12.5 percentage points. Similarly, none of the coefficients for asthma are significant, but the lack of precision does not allow us to rule out very large effects (either positive or negative). For instance, the estimates by age 5 suggest an impact ranging from -4.2 pp to 18.4 pp. Finally, results for obesity and atopic diseases are not significant either but also quite imprecise, and the latter are also very large in size.

Similarly to our results for neonatal outcomes, the bivariate probit estimates are much more precise, but still consistent with the 2SLS, given the large standard errors of the latter. For type 1 diabetes, the coefficient is much smaller than the one from the linear model and is no longer 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). Although by age 10 the estimates are a bit noisy and no longer significant, the results in Figure 3 show that unplanned C-sections significantly increase the probability of an asthma diagnosis for children as young as 2, with the effect being statistically significant up to age 9. These estimates are statistically consistent with the 2SLS ones, which are much noisier. For obesity, the bivariate probit results are precisely estimated at zero for both age 5 (0.001, 95% CI 0.000–0.002) and 10 (0.003, 95% CI 0.000–0.006), but the results in Figure 3 show a detectable effect from age 11, which is consistent with puberty being a vulnerable period for the development of overweight and obesity (Lobstein et al., 2004). Lastly, we do not find a significant impact on atopic diseases at age 5 or 10.

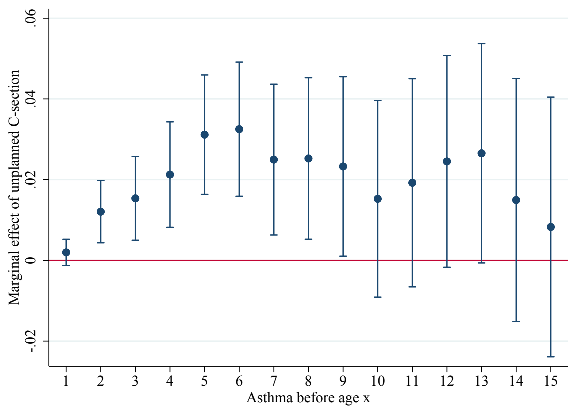
Interestingly, the estimates of the siblings fixed effect model are very similar to the bivariate probit results, both in significance and in magnitude. We only find an increased risk of having an asthma diagnosis by age 5 for second-borns compared to first-borns in families where the second child was born by C-section compared to families where both children were born by vaginal delivery. Similarly to the bivariate probit estimates, Figure 4 shows that this effect is significant from ages 1 to 8. Although, as discussed in Section 3.2, our sibling fixed effect estimates could be negatively biased, we do not find any significant effect on obesity, atopic diseases, or type 1 diabetes, reinforcing the conclusion that C-sections do not have an impact on these outcomes.

Table 3: IV and bivariate probit results – Child diagnoses by age

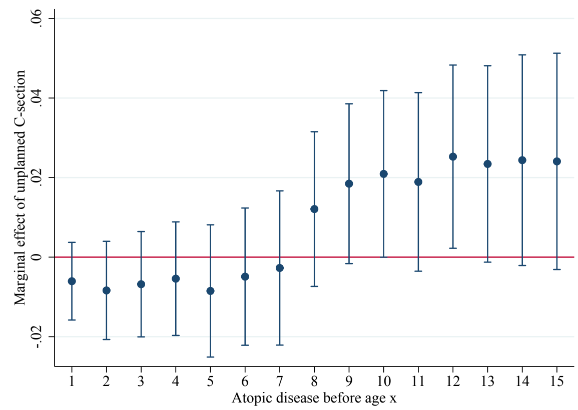
	Type 1 diabetes		Asthma		Obesity		Atopy	
By age:	5	10	5	10	5	10	5	10
OLS	0.000	0.000	0.007***	0.010***	0.000***	0.002***	0.002**	0.004***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
$\overline{Y}$	0.003	0.006	0.045	0.071	0.001	0.004	0.044	0.061
N	807035	556009	807035	556009	807035	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
Siblings f.e.	-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

Notes: 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 siblings fixed effects estimation (see equations (2), (3), and (4)). Specifications as detailed in section 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 siblings F.E. panel. First-stage F statistic from 2SLS and bivariate probit specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$

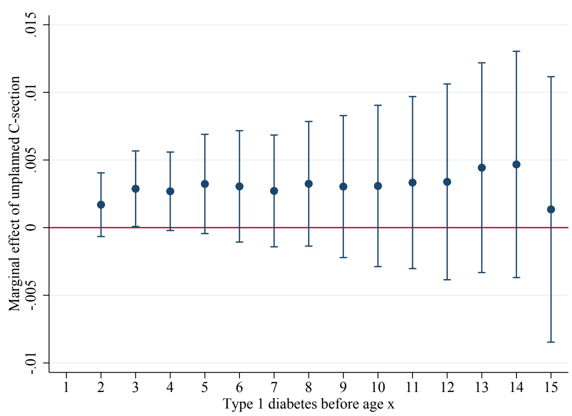
Figure 3: Bivariate probit estimation – Child diagnoses by age



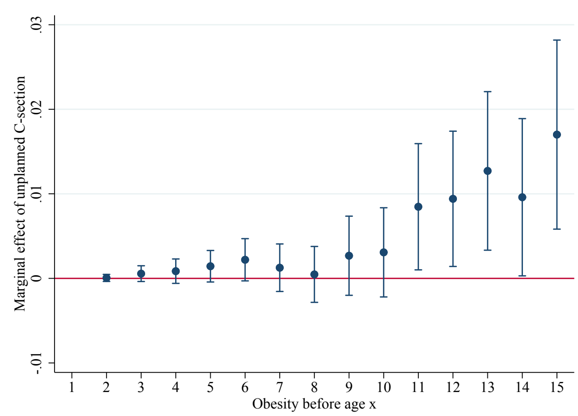
(a) Asthma



(b) Atopy



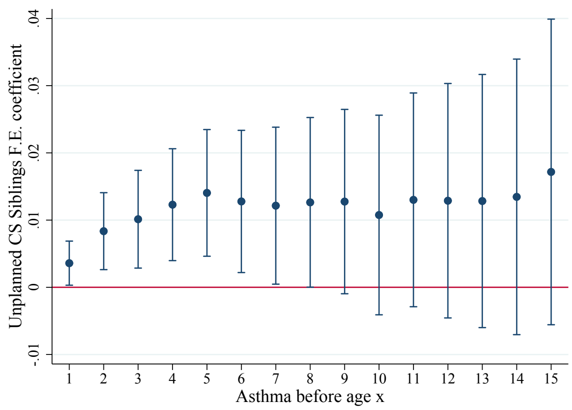
(c) Type 1 Diabetes



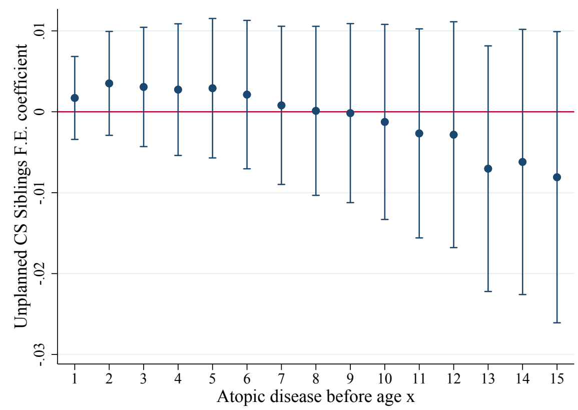
(d) Obesity

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 3.2.1.

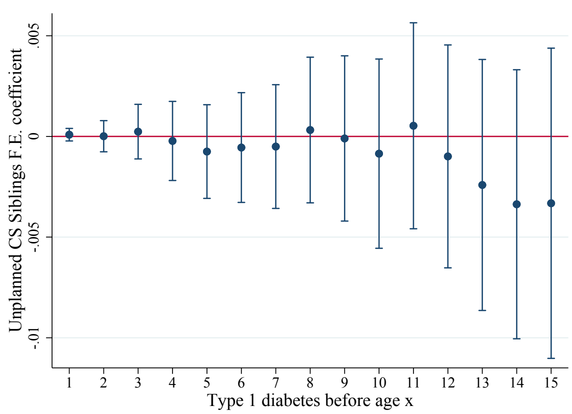
Figure 4: Within-family analysis – Child diagnoses by age



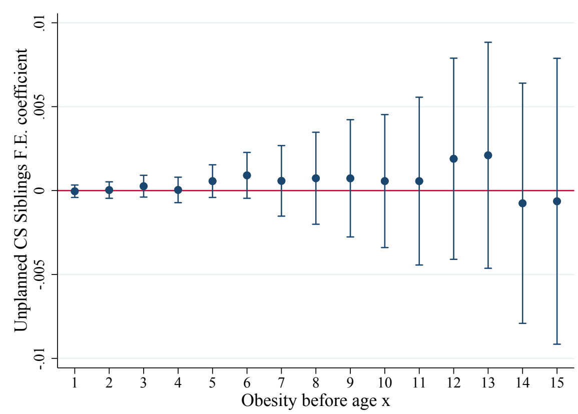
(a) Asthma



(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 3.2.2.

All in all, 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: bivariate probit marginal effects indicate a slightly larger impact, of around 2 pp considering the different ages, but are more imprecise than those from the within-family analysis, which are around 1.2 pp. Comparing these increases to the sample mean, the noisier bivariate probit suggests an increase in the probability of asthma of 50% (compared to the mean of 4%), while the within-family model suggest an increase of 27% (compared to the sample mean of 4.5%). The latter is closer to the reported associations of a 20% increase in the risk of asthma for C-section babies documented in recent meta-analyses (Thavagnanam et al., 2008; Keag et al., 2018).

On the other hand, 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 with the bivariate probit, and larger than 0.1 pp with sibling fixed effects. For atopic diseases, in turn, our results discard effects larger than 1.2-1.3 pp with both methods. Finally, although bivariate probit results suggest there might be an effect of C-sections on obesity after age 11, our analysis is not conclusive in this regard, as the sibling fixed effect results do not corroborate this finding. For younger ages all methods suggest no impact of cesarean birth on obesity: for instance, estimates at age 5 rule out effects larger than 0.3 pp.

A caveat should be kept in mind when interpreting these results: these estimates are only informative of the impact of unplanned C-sections on hospital diagnoses for these conditions. While, for some of them, these diagnoses might be a good approximation to the true prevalence of the disease, for others this might not be the case. For instance, a previous study of type 1 diabetes in Finland documents that, in this country, children with newly diagnosed type 1 diabetes are almost always treated in the hospital, and are therefore listed on the Hospital Discharge Register (Harjutsalo, 2008). This implies that we are able to observe most type 1 diabetes diagnoses in our population of interest. On the other hand, for asthma, diagnoses are made by general practitioners since 1994 in Finland (Tuomisto et al., 2010) and we might thus only be able to trace the most severe cases. The same might be true for atopic disesase and obesity.[24] 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.

5 Validity checks

5.1 Exclusion restriction and se sensitivity checks

In the case of our instrumental variable, our setting makes it very unlikely that the exclusion restriction is violated, that is, that the instrument affects infant health through different channels other than the increased probability of cesarean section. This would require other changes to happen during pre-leisure days and only during the regular shift. In what follows we provide further evidence supporting this assumption.

Given that we are comparing working days, the activity at maternity wards should be equivalent across different types of days. The first panel of Figure A4 shows the proportion of planned cesarean sections by type of day and time of birth, evidencing that scheduled activity is organized very similarly over the day in pre-leisure days compared to other working days. Furthermore, we do not find evidence of maternity wards' crowding during these days, as shown by comparing the number of births by type of day and weekday in the second panel of Figure A4.

Regarding the quality of care provided during these days, the first panel of Figure A5 shows that the probability of having a low Apgar score (below 7) does not differ across weekdays or types of days, suggesting that quality of care during labor and delivery does not differ by type of day. Similarly, panel two of Figure A5 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. This measure could capture changes in quality of care the days after birth. Again, we do not find evidence that early neonatal mortality is higher for babies born on pre-leisure days. Moreover, we do not find differences in mothers' length of stay among those who had a C-

[24]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).

section by weekday or type of day (Figure A6), suggesting that other factors are kept constant across different days.

Since babies born on pre-leisure days stay in the hospital during a public holiday or the weekend, one could argue that the quality of care they receive is worse compared to children born on other working days. In Table 4 we show the coefficients from performing the same IV regressions but restricting the sample to babies born on Thursdays or Fridays. Given that the average stay in our sample is four days, all infants born both on Fridays and Thursdays were hospitalized during the weekend. In spite of the reduced sample size, results from this exercise are consistent with our previous estimates, providing further evidence in favor of our exclusion restriction.

Although in Figure 2 we do not find that mothers who give birth during regular working hours on pre-leisure days have a higher probability of having had their labour induced, these mothers are more likely to fall into a ”gray area” that offers more room for discretionary behaviour,[25] and thus, the decision to perform a C-section in these cases might be more sensitive to doctors' leisure incentives. In other words, these mothers are more likely to be part of the complier population. In column 3 in Table 4 we show that our coefficients remain about the same if we exclude from our sample mothers whose labor was induced. The same is true if we exclude inductions from our siblings fixed effects estimation. This result confirms that our findings are not driven by this sample of mothers only.

Finally, for the case of the siblings fixed-effects model, we run a placebo regression and focus on an outcome that should not be affect by the C-section: birth weight. Results can be found in the first column of Table 4. We do not find that being born by C-section predicts birth weight, suggesting that these babies do not have worse health in general. This result supports the validity of this strategy: family fixed-effects, jointly with the controls, seem to be taking into account general health differences between siblings born by C-section and vaginal delivery.

[25] 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.

Table 4: Validity checks

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

Notes: 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 siblings fixed effect (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 3.2.1 and 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 siblings F.E. panel. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$

5.2 Within-family validity check

As explained in section 3.2.2, in our siblings fixed effects model, although we control for a large set of time-varying observable characteristics, we could expect our estimates to have a negative bias, if any. Even if we do not find effects on the placebo outcome (birth weight), compared to their vaginally-delivered older siblings, second children born by C-section might have had some problems either during pregnancy or during the delivery that we cannot observe in our data. As shown in Section 4, the results from this estimation suggest that C-sections do not have an impact on some diseases that had previously been associated with cesarean births, but we do find that they lead to a positive increase in the probability of having asthma.

In order to assess to what extent these results could be driven by selection, rather than by the C-section itself, we compare these estimates to those from similar samples of sibling pairs where we expect the second child to be negatively selected with respect to their older sibling, but where C-sections do not play a role. In particular, we select two samples: a sample of siblings where the first child was born by eutocic birth, and the second child was born either by eutocic or by instrumented birth;[26] and 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 (but all of them were born by vaginal delivery).[27] We will thus compare the health gap between siblings in families where there was a problem during the second birth (leading to the use of instruments to assist the delivery) or during the second pregnancy, with respect to families where none of the siblings encountered any of these issues during pregnancy or birth.

Table 5: Validity sibling f.e. results – instrumented vs. eutocic and low vs. high-risk pregnancy

	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

Notes: This table shows the results from sibling fixed effect models, following the specification in equation (4), 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 3.2.2. Standard errors are clustered at the family level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$

The results from this exercise can be found in Table 5. The first four columns show that, with

[26]An eutocic delivery is a vaginal delivery with no instrumentation, that is, without ventouse or forceps assistance.

[27]We define a high-risk pregnancy as one in which the mother had at least one of these problems: a positive result in the glucose tolerance test or 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.

respect 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 second siblings who experienced a high-risk pregnancy do not have significantly worse neonatal health by any of the indicators, although all the coefficients are positive. In the last four columns, in turn, we explore if this negative selection is reflected in a higher probability of having any of the diagnoses we analyze in section 4. However, 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. This suggests that, while our estimates of the impact of C-sections on neonatal health from the within-family model could be biased by negative selection of the cesarean born sibling, this selection is unlikely to explain the specific effects we find for asthma.

6 Conclusions

This paper provides new evidence on the effects of avoidable cesarean sections on various infant health outcomes. In order to overcome potential omitted variable bias and abstract from those cases in which C-sections respond to a clear clinical indication, we utilize a novel instrumental variable that gives exogenous variation for unplanned C-sections. Our instrument exploits the finding that unplanned C-sections are more common on regular working hours on Fridays and working days preceding public holidays. We complement this strategy by estimating siblings fixed effect models and comparing the health gap between siblings in families where the second child was born by unplanned C-section with the gap between siblings who were both born by vaginal delivery.

Our results suggest that C-sections have a significant negative impact on neonatal health, although this effect is not severe enough to translate into an increased mortality risk. In our longer run analysis, we follow children from birth to age 15 and investigate the impact of C-sections on four hospital diagnosis that have been consistently associated with C-sections: type

1 diabetes, asthma, obesity, and atopic diseases. In contrast to the OLS results, with our instrumental variable and with the siblings fixed effects model we do not find unplanned C-sections to 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 some previous studies (e.g. Thavagnanam et al., 2008; Keag et al., 2018).

In this paper we provide the first long-run evidence of the causal impact 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, in general, 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 that give birth at regular working times 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, the results of this analysis 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. This paper thus hopes to provide a solid base upon which future research on the effects of avoidable cesarean sections can be built.

References

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

Andrabi, T., Das, J., and Khwaja, A. (2013). Students today, teachers tomorrow: Identifying constraints on the provision of education. *Journal of Public Economics*, 100:1–14. cited By 15.

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 1 edition.

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).

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.

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.

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.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2016). Does grief transfer across generations? bereavements during pregnancy and child outcomes. *American Economic Journal: Applied Economics*, 8(1):193–223.

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.

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

Card, D., Fenizia, A., and Silver, D. (2018). The health effects of cesarean delivery for low-risk first births. NBER Working Papers, 24493.

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

Dominguez-Bello, M. G., Jesus-laboy, K. M. D., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., Bokulich, N. A., Song, S. J., Rivera-vina, 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(3):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.

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

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.

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

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

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.

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.

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.

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.

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.

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.

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.

Li, H., Zhou, Y., and Liu, J. (2013). The impact of cesarean section on offspring overweight and obesity: a systematic review and meta-analysis. *International Journal of Obesity*, 37(7):893–899.

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.

Magne, F., Puchi Silva, A., Carvajal, B., and Gotteland, M. (2017). The Elevated Rate of Cesarean Section and Its Contribution to Non-Communicable Chronic Diseases in Latin America: The Growing Involvement of the Microbiota. *Frontiers in Pediatrics*, 5(September).

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.

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 (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.

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.

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.

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.

Saccone, G. and Berghella, V. (2015). Induction of labor at full term in uncomplicated singleton gestations: a systematic review and metaanalysis 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.

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

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.

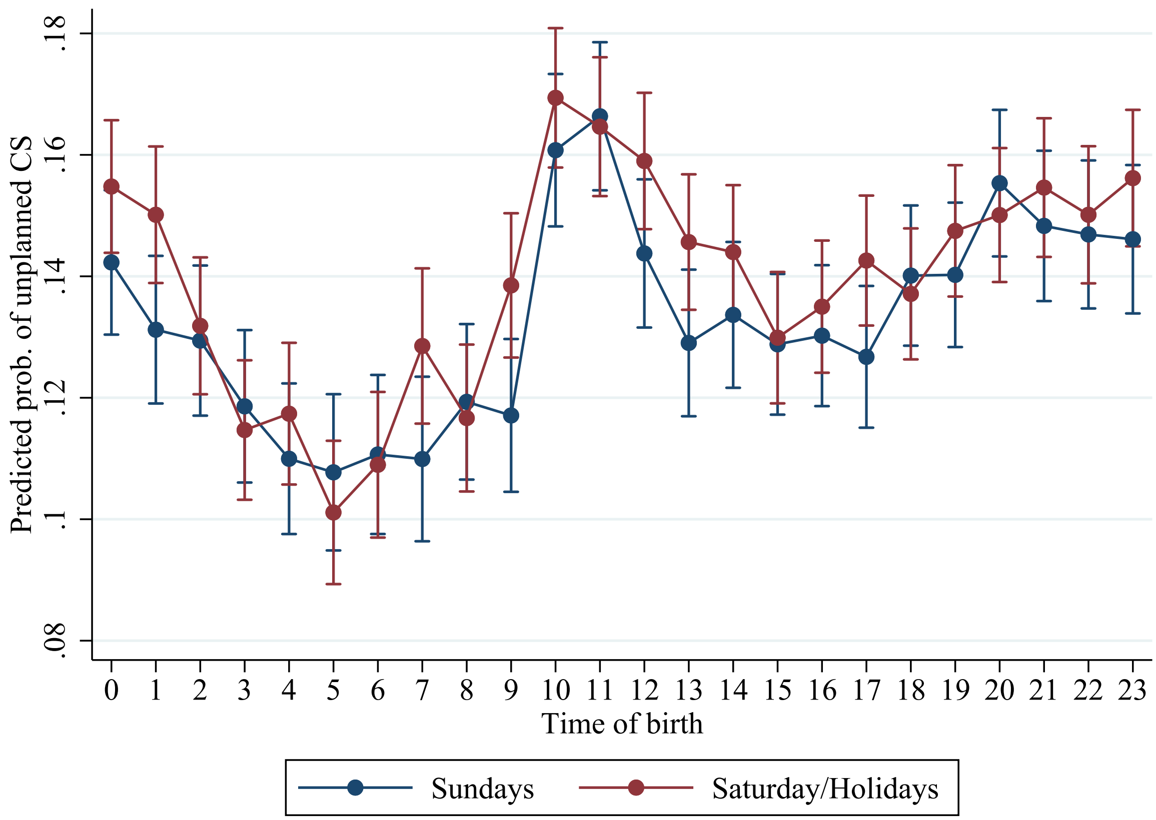
Torpy, J. M. (2010). Chronic diseases of children. *JAMA - Journal of the American Medical Association*, 303(7):682.

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.

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

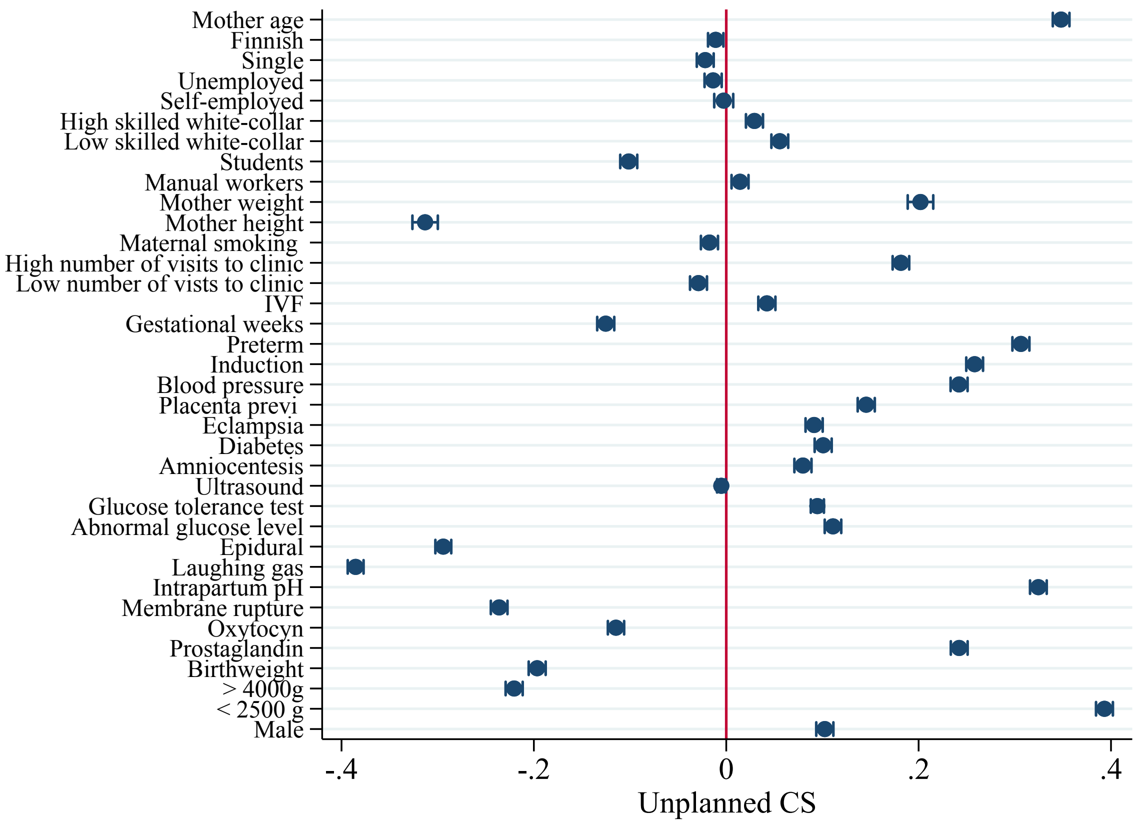
Appendix

Figure A1: Predicted probability of unplanned C-section by time on weekends



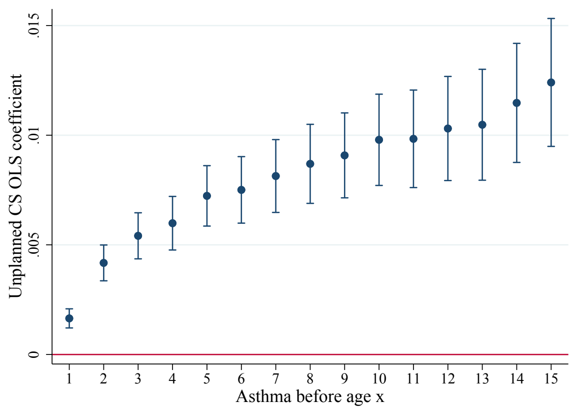
Notes: The figure represents the predicted probability of unplanned C-sections by time of birth for Sundays and for Saturdays or holidays, adjusting for hospital, month, and year of birth fixed effects. Sample is restricted to single births, unscheduled C-sections and vaginal births that take place on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

Figure A2: Difference in baseline characteristics by type of birth

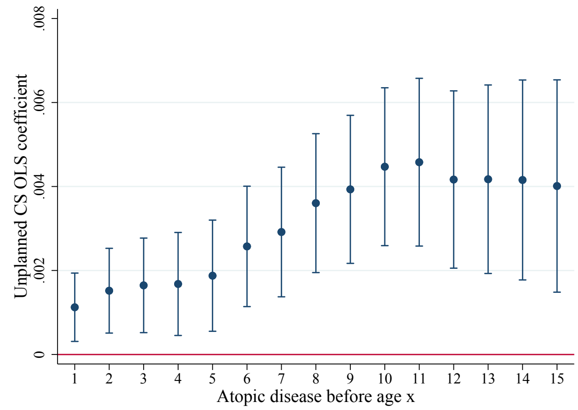


Notes: The figure represents the coefficients and 95% CI from separate regressions of each (standardized) predetermined variable on an indicator taking value 1 if the mother had an unplanned C-section, and 0 if it was a vaginal delivery, controlling for normal shift time, pre-leisure day, and hospital, 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.

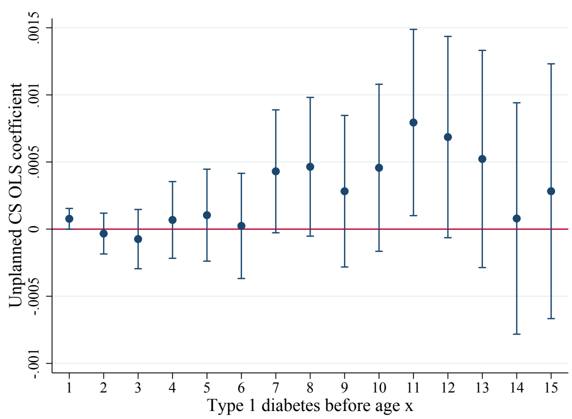
Figure A3: OLS estimation: Child diagnoses by age



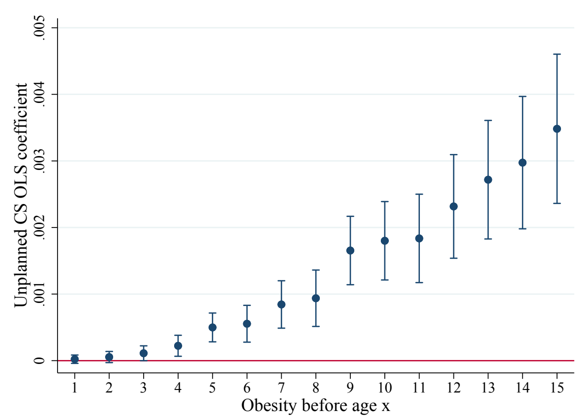
(a) Asthma



(b) Atopy



(c) Type 1 Diabetes

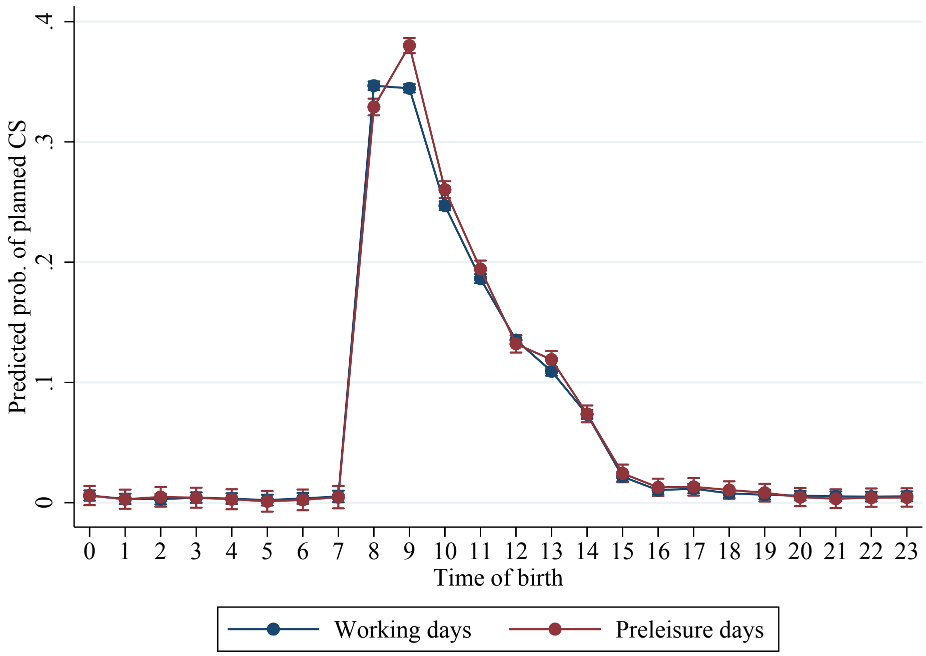


(d) Obesity

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 3.2.

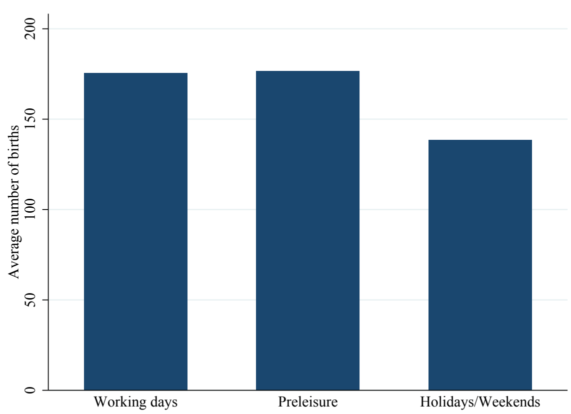
Figure A4: Activity at maternity wards by type of day

Planned C-sections

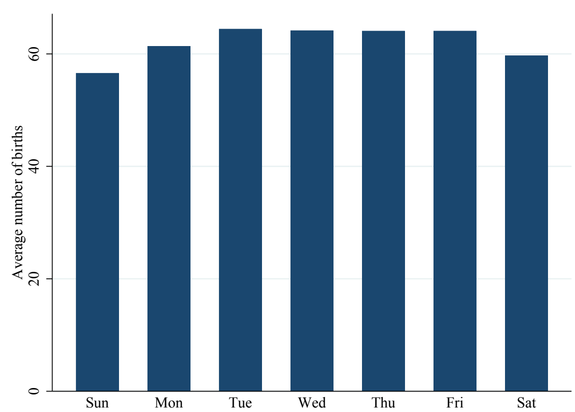


Number of births

(a) By type of day



(b) By weekday

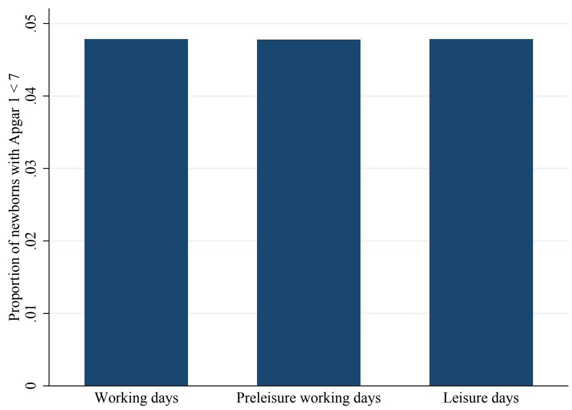


Notes: This figure plots, in the first panel, the probability of planned C-section by time of birth on pre-leisure working days and other working days, and in the second panel, the average number of births by type of day (column (a)) and by weekday (column (b)).

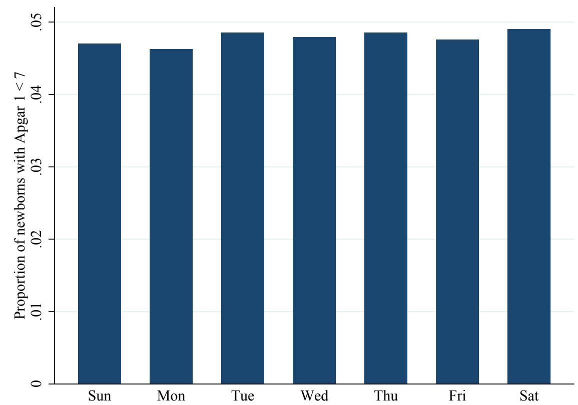
Figure A5: Quality of care by type of day

Low Apgar score

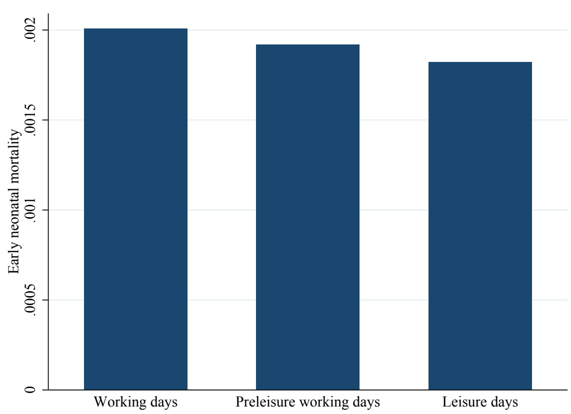
(a) By type of day

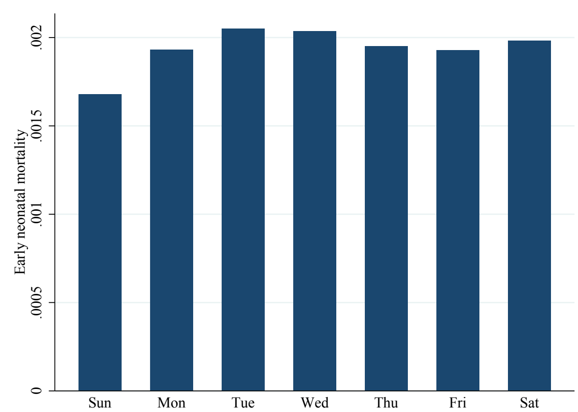


(b) By weekday



Early neonatal mortality

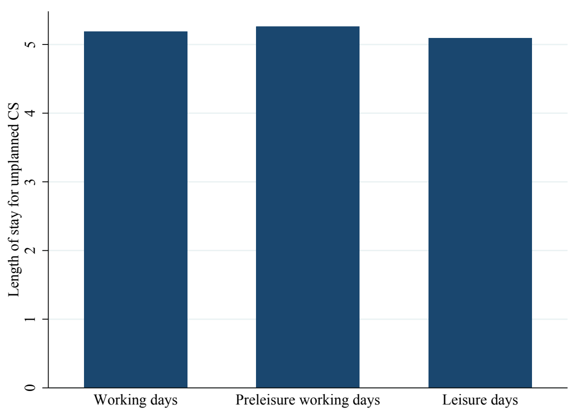




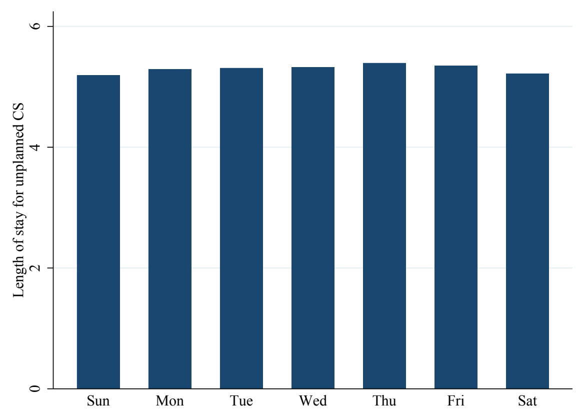
Notes: This figure plots, in the first panel, the probability of the newborn having low apgar score and in the second panel the probability of early neonatal mortality by type of day (column (a)) and by weekday (column (b)). Sample is restricted to single births, unscheduled C-sections and vaginal births.

Figure A6: 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 A1: 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 A2: 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 A3: 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 A4: 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

Notes: 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). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$

Table A5: 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

Notes: 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). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$