The Long-Run Effects of Cesarean Sections

Ana Costa-Ramón

Mika Kortelainen

Ana Rodríguez-González

Lauri Sääksvuori

Abstract

This paper analyzes the long-term effects of potentially avoidable cesarean sections on children's health. Using Finnish administrative data, we document that physicians perform more unplanned C-sections during their regular working hours on days that precede a weekend or public holiday and use this exogenous variation as an instrument for C-sections. We supplement our instrumental variables results with a differences-in-differences estimation strategy that exploits variation in birth mode within sibling pairs and across families. Our results suggest that avoidable unplanned C-sections increase the risk of asthma, but do not affect other immune-mediated disorders previously associated with C-sections.

1

Supplementary materials are available in the Online Appendix.

I. Introduction

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

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

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

The cesarean section is the most commonly performed major surgery in many countries. Understanding the consequences of cesarean sections on later-life health and human capital

3

development is important from a number of perspectives varying from clinical decision making to economic and health policy. The rapidly growing incidence of cesarean sections across the globe suggests that even small increases in mortality and morbidity due to C-sections would lead to large reductions in life expectancy and substantial losses of human welfare.[1]

This paper provides new evidence on the effect of potentially avoidable cesarean sections on several relevant health outcomes. To identify the causal effect and abstract from cases where C-sections respond to a clear medical indication, we exploit variation in medical decision-making depending on the type of day and time of birth. We show that the probability of unscheduled C-section increases substantially during the normal working hours (8am – 4pm) on working days that precede a leisure day. Importantly, we find that these excess C-sections are not driven either by selection of different mothers giving birth at these times or by advancing births that would have been cesarean deliveries in any event.

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

We investigate the effects of cesarean sections on infant and child outcomes using a comprehensive and precise administrative data resource that includes birth and health records for all children born in Finland between 1990 and 2014. We follow entire birth cohorts from birth to

4

teenage years and use detailed diagnosis data to study the causal effects of cesarean sections on children's health. We focus on outcomes whose onset is hypothesized to be influenced by cesarean delivery: asthma and other atopic diseases, type 1 diabetes, and obesity. These are among the most common chronic conditions in childhood (Torpy, 2010).[2] Our instrumental variable estimates suggest that avoidable C-sections increase the probability of asthma diagnosis from early childhood onward. This effect is clinically and economically relevant. However, we do not find consistent evidence that cesarean sections affect the probability of developing atopic diseases at large, type 1 diabetes, or obesity.

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

Our results are consistent with the hypothesis that the mode of delivery may influence the development of the immune system and have long-term effects on health and disease. However, our results paint a more nuanced picture about the long-term effects of cesarean deliveries than existing evidence based mostly on associations. We show that controlling for the observable characteristics that most previous studies have accounted for is not enough to deal with the endogeneity of the type of birth. Our findings suggest that C-sections cause a much narrower spectrum of diseases than currently hypothesized and call for a careful analysis on the relationships between the delivery mode and long-term health.

5

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

Second, we study the effects of discretionary unplanned C-sections that could potentially be avoided, while existing papers have not been able to separate planned (elective) and unplanned C-sections or have concentrated on C-sections with a clear medical indication. Hannah et al. (2000), Jensen and Wüst (2015), and Mühlrad (2017) show that breech babies can benefit from C-section delivery. However, these results concern medically necessary C-sections in a specific high-risk group and do not readily generalize to avoidable unplanned C-sections or cesarean deliveries in general. While C-sections are often life-saving at the top of the risk distribution (Currie and Macleod, 2017), more evidence is required about the effects of discretionary C-sections that could be potentially avoidable.

Third, to evaluate the causal effects of C-sections, we use two different identification strategies based on somewhat different assumptions. Our instrumental variable strategy builds on previous work using time variation in C-section rates in combination with high-quality administrative data. Moreover, we employ a differences-in-differences research design that has not been used in previous papers on C-sections. For both methods we provide several pieces of evidence that support the credibility of the identification assumptions. Thus, by using two different strategies, we hope to provide more reliable evidence on the causal effects of avoidable unscheduled interventions at birth on children both in the short and long run. Furthermore, we reconcile our findings with those from previous associational studies by showing that controlling for even a very rich set of observable characteristics is not enough to deal with the endogeneity of delivery mode.

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

II.   Background

A. Mechanisms

A large body of literature documents the developmental origins of health and disease. The process of labor can be seen as one crucial step in adaptation to the extrauterine environment. The prevailing evidence highlights the role of vaginal delivery as an important early programming event with potentially life-long consequences (Hyde et al., 2012). While there is strong consensus that medically indicated cesarean sections decrease the risk of fetal death at birth, the absence or

modification of vaginal delivery has been linked to several adverse health outcomes and anomalies in human development. In the following, we summarize some of the most widely acknowledged findings to understand how C-sections might have long-lasting effects on health and human development.

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

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

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

B. Classification of Cesarean Sections

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

Most C-sections are not scheduled and happen after spontaneous or medically induced onset of labor. Unscheduled C-sections are surgeries where an attempt of vaginal birth is transformed to a cesarean delivery after the mother has been admitted to a hospital. Unscheduled C-sections are classified by urgency. Emergency C-sections are performed within 30 minutes of the decision, due to an immediate threat to the life of the mother or the baby (NICE, 2011). However, most unscheduled C-sections are performed without such immediate threat. The optimal timing and indication for these operations are imprecise and give large discretion to the clinician. Slow

progression of labor or cephalopelvic disproportion are examples of diagnoses that may require an unplanned non-urgent cesarean section. There is wide variation among clinicians in the use of discretionary diagnoses that justify C-sections (Barber et al., 2011; Fraser et al., 1987). Our data contains the registered diagnosis linked to the C-section for a subsample of births. These observations enable us to verify that the peaks in unplanned C-sections are coupled with the use of more discretionary diagnoses.

C. Institutional Context

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

Pregnant women usually give birth in the nearest hospital. Only high-risk pregnancies are systematically directed to a higher-level hospital for obstetric care and delivery. Expectant women do not have pre-assigned midwifes or physicians for the delivery. Midwives take care of the delivery in all hospitals, while physicians have the ultimate responsibility for obstetric care, decide on the type of delivery, and perform C-sections. Physicians' financial incentives to perform a C-section are negligible in Finland. The total income of doctors working in public hospitals consists of basic salary (63%), on-call and overtime payments (28%), personal supplements (e.g. years of service) (7%), and performance-based pay components (e.g. fees per appointment or procedure) (3%) (Finnish Medical Association, 2016). The C-section rate (15.5% in 2015) is relatively low from an international perspective (OECD, 2017).

The regular working shifts for physicians are from 8am to 4pm from Monday to Friday. The on-call hours for physicians may not exceed 24 hours during the regular working week and last

typically from 8am to 8am. On weekends, the on-call hours for physicians are from 8am to 9am on the next day.[8] Midwives follow the same rotation regardless of the type of day and work in three 8-hour shifts.[9]

While our data does not contain details about the organization of hospital resources across different shifts, we have collected data on the aggregate number of midwives and physicians by weekday and work shift for three different hospitals: two university hospitals and one small hospital. The qualitative evidence from these three hospitals, described in Table A1 in the Online Appendix, shows that staff availability does not vary between working days.

III. Data and methods

A. Data

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

We exclude from our sample planned C-sections and multiple pregnancies. For our instrumental variable strategy, we focus only on first births.[11] Our analysis sample includes 392,560 deliveries that took place from 1990 to 2014. For the differences-in-differences analysis, we focus on both first and second births from families where the first child was born by vaginal delivery (more

details are provided in Section III.B.2). The analysis sample consists of 645,292 children from 322,646 sibling pairs. There are 43 hospitals in our sample. Table A2 in the Online Appendix shows summary statistics for all births in Finland between 1990 and 2014.

We match the Finnish Medical Birth Register to the Finnish Hospital Discharge Register, which contains information about the diagnosed medical conditions, medical operations, and the date of diagnoses. This hospital register contains all inpatient consultations in Finland from 1990 to 2013. From 1998, the data include all outpatient visits to hospitals. All diagnoses are coded using the International Classification of Diseases (ICD) tool.[12]

We explore two sets of outcome variables. First, to test whether unplanned C-sections have an impact on neonatal health, we analyze indicators of neonatal health included in the birth register. We study Apgar scores one minute after birth, admission to intensive care unit (ICU), need of assisted ventilation, and early neonatal mortality (defined as neonatal death in the first week of life).[13] Second, we study longer term outcomes using detailed inpatient and outpatient diagnosis data from the Finnish Hospital Discharge Register. We use primary diagnoses.[14] 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 A3 in the Online Appendix provides more detail about each of these diagnoses.

B. Empirical strategy

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

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

where Y_i is the health outcome of infant i, X_i is a vector of covariates and δ_m, λ_y, ϕ_h are fixed effects for the month, year, and hospital of birth, respectively.[15]

The estimation of Equation (1) is, however, likely to provide biased estimates of β_1 due to potential selection into cesarean birth.[16] To study the causal effects of cesarean delivery on children's health, we exploit two different empirical strategies.

1. IV strategy: Variation by time and type of day

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

Figure 1 presents the predicted probability of 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 C-sections over a 24-hour cycle for working days that precede a leisure day compared to other working days.[17] We find that substantially more C-sections are performed during regular working hours on days that precede a leisure day compared to the rest of working days. Figure 1b presents the predicted probability of having an unplanned C-section by work shift and type of day. We find that the gap in C-section rates between a day that precedes a leisure day and the rest of working days emerges only during the regular working hours (from 8am to 4pm).[18]

Importantly, we find that the excess C-sections performed in days that precede a leisure day are not driven by advancing births that would have been cesarean deliveries in any event. We do not observe any relative fall in C-sections during the evening hours preceding a leisure day compared to the evenings of regular working days (Figure 1a) or during the leisure day (Figure A2 in the Online Appendix).[19] These observations suggest that physicians perform C-sections during the regular working hours on pre-leisure days that would not have been performed otherwise.

[Figure 1]

The time pattern of C-sections is consistent with previous work by Brown (1996) and Halla et al. (2020) that documents an increase in C-section rates on days that precede a leisure day. Halla et al. (2020) exploit this variation in an instrumental variable framework to study the impact of delivery mode on maternal fertility and labor supply. The existing literature attributes the pre-leisure anomaly in the time pattern of C-sections to physicians' leisure incentives that arise from the higher time cost and uncertainty of vaginal births. A cesarean section takes on average 30-75 minutes and is perceived as a relatively easy surgical intervention with low complication rates (NICE, 2011). The average duration of labor for first-time mothers who have a vaginal birth is 11 hours (NICE, 2014).

We provide two pieces of complementary evidence to validate that the excess rate of C-sections is not driven by medical factors. First, we build on previous evidence that some medical diagnoses linked to a cesarean birth are more discretionary than others. Dystocia (prolonged or obstructed labor), one of the most common indications for primary cesarean section, is believed to provide the greatest room for diagnostic discretion (Fraser et al., 1987). The number of dystocia diagnoses has been shown to strongly respond to physician incentives (Evans et al., 1984; Fraser et al., 1987; McCloskey, Petitti and Hobel, 1992). We examine if there is an excess number of dystocia diagnoses during regular working hours on pre-leisure days. Our results (Table A5 in the Online Appendix) show that giving birth during the regular hours on a pre-leisure day increases the probability of having a dystocia diagnosis compared to other working days. Importantly, we do not find this temporal pattern for medical emergencies, for which there should not be any room for discretion. In particular, we find that our instrument does not predict additional examinations of the fetus during labor, which doctors should perform if there are any signs of fetal suffering.[20]

Our second piece of evidence builds on the literature showing that physician mothers are less likely to receive C-sections driven by financial incentives (Johnson and Rehavi, 2016). Consequently, we expect that the change in medical criteria during the normal shift on pre-leisure days would not affect physician mothers and other medical professionals. Our results (Table A6 in the Online Appendix) support this hypothesis. We do not find that medical professionals have an increased risk of having a C-section during the regular shift on pre-leisure days, while we do find this increase for mothers with an equivalent level of education who are not employed in a medical profession.[21]

Overall, these pieces of evidence suggest that the observed increase in C-section rates is supply-driven and are consistent with the interpretation of previous studies which have attributed similar increases to physicians' leisure incentives (Brown, 1996; Halla et al., 2020).

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

$$CS_i = \gamma_0 + \gamma_1 NS_i + \gamma_2 Preleisure_i + \gamma_3 NS_i \times Preleisure_i$$

$$+ X_i'\gamma_4 + \delta_m + \lambda_y + \phi_h + \upsilon_i \tag{2}$$

and the corresponding second stage:

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

where NS_i is a dummy that takes value 1 for births that take place during the normal shift (from 8am to 4pm) 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 predicted C-sections from the first stage, X_i is the vector of individual controls,[22] and δ_m, λ_y, ϕ_h are month,

year, and hospital of birth fixed effects, respectively. The interaction between regular working hours and a day preceding a leisure day will serve as an instrument. As a result, we will be comparing mothers who give birth in the same hospital during the same shift, but on different types of days (working days preceding a leisure day or other working days).

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

Table 1 shows the results from the estimation of the first stage. Column (1) shows the first stage estimates including month, year, and hospital fixed effects. Column (2) includes a richer set of controls. These estimates show that being born during the normal shift increases the probability of unplanned C-section for all working days. Moreover, being born during the normal shift on pre-leisure days increases the probability of unplanned C-section by 1.4 percentage points.[23] This implies a 9.6% increase with respect to the average unplanned C-section rate of 14.5% in our sample. This estimate is in line with, but larger than that by Halla et al. (2020), which document a 3.6% increase in C-sections on pre-leisure days taking into account both planned and unplanned cesareans. The first stage F-statistics are larger than 25 in both specifications. Following the common critical values for weak instruments (Stock and Yogo, 2005), we can reject the null hypothesis that the instrument is weak.

Figure 2 shows that our instrument does not predict a large set of maternal and pregnancy characteristics, including medical conditions that could predict a C-section. This indicates that mothers giving birth during the regular shift on pre-leisure days compared to other working days

are similar in observable characteristics, suggesting that the observed increase in C-sections at these times cannot be explained by selection.[24]

[Table 1]

Finally, regarding the exclusion restriction, we focus on births that take place on working days, when hospital resources and quality of care should be constant. Moreover, to compromise our empirical strategy, any change in the quality of care would need to happen on pre-leisure days only during the regular working hours. The information about the organization of hospital staff discussed in Section II.C (Table A1 in Online Appendix) suggests that this kind of change is unlikely in our setting, as there is no evidence of differences in the distribution of hospital resources across shifts between different types of working days.[25] In Section V.A, we show that the planned activity in the maternity wards during the regular shift is very similar on pre-leisure days compared to the rest of working days, and provide numerous supplementary analyses that further reinforce the credibility of the exclusion restriction.

The two-stage least squares estimator enables us to identify a local average treatment effect (LATE). This is the effect of C-sections for infants whose mothers' mode of delivery is sensitive to the subjective assessment of the physician. More accurately, we capture births where the type of day affects the decision of the doctor to perform a C-section during the normal shift. The counterfactual for these births is unlikely to be exclusively a cesarean section later on, given that we do not find a relative drop in C-sections on pre-leisure days after the normal shift or during the following day.

[Figure 2]

The LATE will not be informative of the effect of medically indicated C-sections, as those will be performed regardless of physician's incentives. Moreover, the LATE does not capture the effect of unplanned C-sections for babies who had a very fast delivery, leaving no room for physician discretion.

Our primary health outcomes and the endogenous variable are binary. Consequently, besides the 2SLS models we estimate (recursive) bivariate probit models. These specifications mirror Equations (2) and (3) and assume that cesarean delivery (CS_i) and the binary indicator of health Y_i are determined by the following latent indices:

$$CS_i = \mathbb{1} \left[\rho_1 NS_i + \rho_2 Preleisure_i + \rho_3 NS_i * Preleisure_i + X_i' \rho_4 + \delta_m + \lambda_y + \phi_h + v_i > 0 \right] \quad (4)$$

$$Y_i = \mathbb{1} \left[\pi_1 NS_i + \pi_2 Preleisure_i + \pi_3 CS_i + X_i' \pi_4 + \delta_m + \lambda_y + \phi_h + \xi_i > 0 \right] \quad (5)$$

where (v_i, ξ_i) follow a bivariate standard normal distribution with unknown correlation. These equations can be estimated through maximum likelihood. Identification in this setting relies on the same assumptions that are needed to estimate the 2SLS model together with an additional assumption about the joint normality of the error terms. In the results section we report marginal effects for both estimators.[26] The first-stage results of the bivariate probit estimation are presented in Table A8 in the Online Appendix and show very similar marginal effects to those from the linear model in Table 1.

Bivariate probit estimation is expected to present substantial advantages in the context of this paper, as it has been shown to be more efficient and less biased than 2SLS when treatment and outcome probabilities are close to 0 or 1 (Chiburis, Das and Lokshin, 2012; Bhattacharya, Goldman and McCaffrey, 2006; Nielsen, Smith and Çelikaksoy, 2009). Given that we work in a low C-section rate setting and examine relatively rare outcomes, we expect bivariate probit to outperform 2SLS in terms of efficiency. In the Online Appendix we show, based on Monte Carlo

simulations for our particular context, that this in indeed the case. The results from this exercise reveal that, with a C-section rate around 15% and having diseases with a prevalence equal or lower than 5% as outcomes, 2SLS results are largely uninformative due to their lack of precision. Bivariate probit results, in turn, are much more efficient, while being comparable to 2SLS in terms of unbiasedness.

2. Differences-in-differences

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

We estimate the following equation:

$$Y_{if} = \psi_0 + \psi_1 Secondborn_{if} + \psi_2 Secondborn_{if} \times CS_{if} + X'_{if}\psi_3 + \gamma_f + \delta_m + \lambda_y + \phi_h + \eta_{if} \quad (6)$$

where Y_{if} is the health outcome of child i in family f, $Secondborn_{if}$ is a dummy variable equal to 1 for the second child and 0 for the first child, CS_{if} is an indicator equal to 1 for unplanned C-section and 0 for vaginal delivery, X_{if} is a vector with the same pregnancy and maternal controls of Equation (3), except for maternal characteristics that are time-invariant, and diagnoses during pregnancy and delivery (like prolonged and obstructed labor),[27] γ_f, δ_m, λ_y and ϕ_h are family, month, year, and hospital of birth fixed effects, respectively.[28] We cluster standard errors at the

family level. Our parameter of interest is ψ2, 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.

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 vaginal birth in the second delivery are a very selected sample, given the very high probability of having a repeat C-section.[29] Second, some studies find that having a C-section is associated with lower fertility (Halla et al., 2020; Keag, Norman and Stock, 2018). We abstract from these concerns by focusing on mothers whose first birth was a vaginal delivery.

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

IV. Results

A. Neonatal outcomes

We first estimate the impact of C-sections on neonatal outcomes. We study 1-minute Apgar scores, admission to the intensive care unit (ICU), assisted ventilation, and neonatal (7- day) mortality. Table 2 shows our OLS (first panel), 2SLS (second panel), bivariate probit marginal effects (third

panel) and differences-in-differences (fourth panel) estimates. For each estimation method, we report in square brackets Romano-Wolf p-values adjusted for multiple hypothesis testing (Clarke, Romano and Wolf, 2019).[30]

We find that the OLS results replicate existing findings. Cesarean sections are associated with adverse outcomes at birth and higher neonatal mortality.[31] Our 2SLS estimates are not significant for any of the outcomes. However, the magnitude of coefficients and large standard errors suggest that we cannot reject that there is a (potentially large) effect on neonatal outcomes. As discussed in Section III.B.1, 2SLS estimates are expected to be particularly uninformative with low treatment and outcome probabilities.

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

[Table 2]

B. Later child health

We now turn to the long-run effects of C-sections on health outcomes. Table 3 shows the OLS (first panel), two-stage least squares (second panel), bivariate probit (third panel) and differences-in-differences (fourth panel) marginal effect estimates at ages 5 and 10. We analyze health conditions that have been extensively documented in the literature as being positively associated with cesarean deliveries: type 1 diabetes, obesity, asthma, and other atopic diseases (atopic dermatitis and allergic rhinitis). Romano-Wolf p-values adjusted for multiple hypothesis testing are reported in square brackets (Clarke, Romano and Wolf, 2019). We show year-by-year bivariate probit and diff-in-diff estimates up to age 15 in Figures 4 and 5, respectively. Our OLS estimates can be found in Figure A4 in the Online Appendix. Given that we study health outcomes for children who are born from 1990 to 2014, the sample size decreases as we consider older ages.

1. OLS results

The OLS estimates (Table 3 and Figure A4) suggest that cesarean sections are associated with a higher probability of asthma, obesity, and atopic diseases. These findings are consistent with existing studies that have documented significant associations between cesarean sections and metabolic and immune-related conditions. However, we do not detect that C-sections are associated with a higher probability of a type 1 diabetes diagnosis.

To compare our findings with the results from previous studies, we repeat in Figure 3 the OLS estimation for the probability of having each disease by age 5 and show how the coefficient changes as we vary the set of included controls. We reviewed the literature on the relation between cesarean delivery and these diseases and made for each outcome a list of the most common control variables.[32] In these regressions, we include all C-sections (planned and unplanned) given that most studies are not able to control for this.[33]

[Table 3]

The first coefficient in each panel of Figure 3 shows the estimate when no additional covariates are included in the regression. We then start to cumulatively add sets of controls. First, we add those included in at least 50% of the reviewed papers, followed by those included in at least 20% of them (see Table A9 in the Online Appendix for a detailed list). In the next specification, we add the rest of the variables included in our usual group of controls (described in Section III.B) that previous literature was not accounting for.[34] Finally, in the last five columns we sequentially add hospital fixed effects, year of birth fixed effects, month of birth fixed effects, we restrict the analysis to unplanned C-sections and, finally, include family fixed effects.

The pattern of this figure points out that controlling for a richer set of observables, as well as including fixed effects, does not significantly change the magnitude and sign of the coefficients. Three types of controls seem to have, however, a somewhat larger effect on the estimates. First, we see for most outcomes that estimates become smaller when controlling for the more extensive list of covariates included in previous papers (the 20% list) compared to controlling for the shorter list of controls included in 50% of them. A key difference between these two groups is that the 20% list includes controls for whether the mother had been diagnosed, before the child's birth, with that particular disease. The second substantial difference in estimates arises when we restrict the sample to exclude scheduled C-sections. Finally, estimates also become closer to zero when family fixed effects, which control for unobserved family heterogeneity, are included.

In any case, even with the richest set of controls, our main conclusion holds: C-sections are strongly associated with an increased risk of being diagnosed with asthma, obesity, and atopic diseases. Controlling for observable characteristics is not enough to alter this finding. For type 1 diabetes we only see a significant association when the smallest list of controls is included. This

association vanishes as we include a stricter set of controls and fixed effects. This reconciles the findings from our OLS model in Figure A4 with the results from previous literature.

2. IV results

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

Similarly to our results for neonatal outcomes, the bivariate probit estimates (marginal effects) are substantially more precisely estimated than the 2SLS coefficients. Yet, practically all point estimates from the bivariate probit models are within the confidence intervals of the 2SLS estimates. For type 1 diabetes, the marginal effect is much smaller than the coefficient from the linear model and not significant. For asthma, the results suggest a significant increase in the probability of a diagnosis by age 5 of 0.031 (95% CI 0.022–0.04). Even though estimates are noisier and no longer significant by age 10, the results in Figure 4 show that unplanned C-sections significantly increase the probability of an asthma diagnosis for children as young as 2 years old. The effect is statistically significant up to age 9. For obesity, the bivariate probit results are precisely estimated at zero at age 5 (0.001, 95% CI 0.000–0.002) and 10 (0.003, 95% CI 0.000–0.006). However, the results in Figure 4 show a statistically detectable effect from age 11. Finally, we do not find a significant impact on atopic diseases at age 5 or 10.

[Figure 3]

[Figure 4]

3. Differences-in-differences results

The differences-in-differences results are very similar to the bivariate probit results. We find that the second-born child has substantially greater risk of having an asthma diagnosis by age 5 than the first-born child in families where the second child is born by C-section. Similarly to the bivariate probit estimates, Figure 5 shows that this effect is significant from ages 1 to 8. Despite the fact that our differences-in-differences estimates could be upwardly biased (Section III.B.2), we do not find any significant effects on obesity, atopic diseases, or type 1 diabetes. These results reinforce the conclusion that C-sections do not have impact on these outcomes.

[Figure 5]

4. Overview of results

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

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

One potential limitation of our analysis is that we study diagnoses made at inpatient or outpatient visits to a hospital. For some outcomes, these diagnoses may be a good approximation to the true prevalence of the disease, while for other diseases hospital diagnoses may lead to underestimation. A previous study documents that in Finland practically all new type 1 diabetes diagnoses are made in a hospital and listed in the Hospital Discharge Register (Harjutsalo, 2008). This evidence implies that we are able to observe practically all type 1 diabetes diagnoses in our population of interest. However, since 1994, diagnoses for asthma in Finland are often made by general practitioners (Tuomisto et al., 2010). Thus, we are likely to trace only the most severe cases of asthma. The same might be true for atopic disease and obesity.[35] In any case, an OLS estimation with the control variables included in most existing studies, or even with a richer set of controls, still yields a significant association of cesarean birth with these hospital diagnoses.

V. Validity checks

A. Exclusion restriction and sensitivity checks

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

First, we show that, conditional on the type of birth, being born during the regular working hours on pre-leisure days does not have any significant correlation with health outcomes. Figure A6 in the Online Appendix shows the reduced-form relationship between having an asthma diagnosis at different ages and the instrument, conditional on being born by C-section (in the first panel) or being born by vaginal delivery (in the second one). The fact that there is no significant reduced-form effect within type of birth is consistent with the instrument being related to child health only through the increased probability of C-section.

Second, we explore the overall activity at maternity wards across the different types of days. The first panel of Figure A7 in the Online Appendix shows the proportion of planned cesarean sections by time of birth and type of day. We find that scheduled activity is organized very similarly during all working days. The same can be seen in the first column of Table 4, where we regress an indicator for planned activity (equal to one for planned C-section or induction) on our instrument, and do not find evidence of any significant difference. Moreover, we compare the number of births by type of day and weekday (Figure A7, second panel) and do not find any evidence of maternity ward crowding during the days that precede a public holiday.

Third, we explore the quality of care provided during different weekdays. The first panel of Figure A8 in the Online Appendix shows that the probability of having a low Apgar score (below 7) does not differ between weekdays or type of day, suggesting that the quality of care during labor

and delivery does not differ by type of day. The second panel of Figure A8 shows the probability of early neonatal mortality, defined as death of a live-born baby within the first seven days of life, by weekday and type of day. We expect that this measure would capture changes in the quality of care after birth. We do not find evidence that early neonatal mortality is higher for babies born on days that precede a public holiday compared to other weekdays. We further explore this issue in columns 2 and 3 in Table 4, where we analyze if during the normal shift on pre-leisure days the likelihood of suffering a medical negligence during childbirth is higher. To proxy for this, we analyze if there is an increased risk of shoulder dystocia or brachial plexus injury.[36] We do not find any evidence that our instrument predicts either of these outcomes, suggesting that the likelihood of medical negligence is not higher during the regular shift on these types of days.

Additionally, we do not find that mothers who have a C-section during the normal working hours on a day that precedes a public holiday have a longer length of stay than mothers who have a C-section at other times. We explore this in column 4 in Table 4, where we regress mother's length of stay on the instrument for the sample of mothers who delivered by C-section. Finally, we explore in column 5 if our instrument predicts a higher risk of postpartum complications for the mother. This is an indicator equal to one if, within one week after the delivery, the mother is diagnosed with any complication related to the puerperium.[37] We do not find any significant difference in this risk on pre-leisure days in general and, during the normal shift in particular, the results suggest if anything a reduction in complications. We interpret all these findings as evidence that the quality of care remains constant across different types of working days during the normal shift.

[Table 4]

Fourth, since babies born on days that precede a public holiday or weekend stay in the hospital during the following non-working days, one could argue that their quality of post-natal care is

worse compared to children born on other working days. This would be constant for both babies born during the regular shift and at other times, and hence would not necessarily compromise the exclusion restriction. Yet, in what follows we assess this concern. Table 5 shows the marginal effects from bivariate probit regressions that restrict the sample to babies born on Thursdays or Fridays.[38] We find, despite the reduced sample size, that the results from this estimation are consistent with our main results.

[Table 5]

Finally, we report in Figure 2 that mothers who give birth during the regular working hours on days that precede a public holiday do not have higher probability of having induced labor. However, the induction of labor is likely to offer more room for discretionary behavior, in which case the decision to perform a C-section might be more sensitive to physician's subjective assessment.[39] In other words, we expect that mothers whose labor has been artificially induced are more likely to be part of the complier population. Column 3 in Table 5 shows that our coefficients remain about the same if we exclude mothers whose labor was induced from our sample. The same conclusion holds if we exclude inductions from our differences-in-differences estimation. These results suggest that our findings are not driven by mothers whose labor has been induced after an admission to the maternity ward.

B. Differences-in-differences validity checks

The results from our differences-in-differences model with family fixed effects could be biased if there are unobservable characteristics correlated with the mode of delivery that vary within family and across siblings. Under this scenario, this methodology would yield upward biased estimates. However, as shown in IV.B.3, our differences-in-differences results suggest that C-sections do not

increase the risk of developing various immune-mediated diseases that have previously been associated with cesarean births.

To assess the extent to which our results could be influenced by selection, we first run a regression using birth weight as a placebo outcome, given that it cannot be affected by unplanned C-sections. Table 5 shows that our differences-in-differences model with family fixed effects does not predict birth weight. This result supports the validity of this strategy: family fixed effects, jointly with the large set of controls, seem to be taking into account general health differences between siblings born by C-section and vaginal delivery.

Second, we compare our differences-in-differences estimates to those from other samples of sibling pairs where we expect the second child to be negatively selected with respect to their older sibling, but where none of them was born by C-section. These samples include: (i) a sample of siblings where the first child is born by eutocic birth and the second child is born either by eutocic or by instrumented birth, and (ii) a sample of siblings where the first born had a low-risk pregnancy and the second born had either a low- or a high-risk pregnancy, while all children in the sample were born by vaginal delivery.[40] Consequently, we assess the health gap between siblings across families that had a complication during the second birth or during the second pregnancy, compared to families where none of the siblings encountered any of these complications during pregnancy or birth.

[Table 6]

Table 6 shows our differences-in-differences estimates using these samples of siblings. The first four columns show that, compared to families where both siblings were born by eutocic birth, second children born by instrumented vaginal delivery have worse neonatal health than their older siblings who had an eutocic birth. We find a significantly higher probability of having low Apgar scores and of being admitted to the ICU (top panel). In the bottom panel, we can see that children

who experienced a high-risk pregnancy do not have significantly worse neonatal health by any of the indicators, even though all coefficients have a positive sign. In the last four columns, we explore if negative selection leading to instrumented birth or risk pregnancy is associated with a higher probability of having any of the diagnoses we analyze in Section IV. We do not find evidence that siblings born by instrumented vaginal delivery or those who had a high-risk pregnancy have an increased risk of type 1 diabetes, asthma, atopic diseases, or obesity at age 5. These observations suggest that our differences-in-differences results for asthma are unlikely to be explained by negative selection.

C. Placebo regressions

In this section, we perform a falsification test and compare the performance of the OLS estimator with that of our instrumental variable strategy and differences-in-differences model. In particular, we analyze, as a long-term placebo outcome, the probability that the child has a diagnosis related to an injury, poisoning or other consequences of external causes at different ages.[41] Results of this exercise can be found in Figure A9. Using OLS, we find that C-sections are associated with an increased risk of the child suffering a hospital admission due to external causes. However, as we would expect, we do not find a causal relationship between C-sections and these diagnoses with either our IV or our differences-in-differences approach. This result reinforces the validity of both approaches to deal with omitted variable bias when examining the impact of C-sections on children's health.

VI.   Conclusions

This paper provides new evidence on the effects of avoidable cesarean sections on various short- and long-term health outcomes. We use a novel instrumental variable estimation strategy to overcome the potential endogeneity of birth mode and abstract from cases in which C-sections

respond to a clear clinical indication. Our empirical strategy builds on the finding that unplanned C-sections are more common during regular working hours on Fridays and working days preceding public holidays. We complement this empirical strategy by estimating a differences-in-differences model with family fixed effects that compares the health gap between siblings in families where the second child was born by unplanned C-section with the health gap between siblings who were both born by vaginal delivery.

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

Our results are consistent with the hypothesis that mode of delivery can affect the development of immune-related conditions, but suggest more nuanced effects of C-sections than previous work. We try to reconcile our results with those in the literature and we find that controlling for the observable characteristics that most previous studies were including is not enough to deal with the endogeneity of birth mode. We illustrate the importance of such endogeneity by performing a placebo test and showing that, even with a rich set of controls, OLS results suggest that children born by C-section are more likely to suffer unrelated health problems, like external injuries. Both

of our identification strategies, in turn, survive this falsification check and suggest that there is no causal effect of cesarean sections on this placebo outcome.

This paper provides novel evidence on the long-term effects of unplanned C-sections that do not respond to a clear medical indication, using inpatient and outpatient data for all children born in Finland from 1990 to 2014. Although we are able to observe most of the cases of type 1 diabetes, for some diagnoses (asthma, atopic disease, and obesity) we might be only able to trace the most severe cases, given that these conditions are often treated by general practitioners. Future work should focus on analyzing the impact of C-sections on obesity and other metabolic disorders using primary care data and anthropometric measurements.

We make use of the detailed diagnosis data to show that variation by time and type of day can be a valid source of variation to investigate the impact of avoidable C-sections. First, we show that mothers who give birth at regular working hours on pre-leisure days are comparable in terms of an 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 work in a medical profession, and whose mode of delivery has been shown by the literature not to respond to doctors' incentives (Johnson and Rehavi, 2016).

All in all, our results suggest that the additional C-sections performed during regular working hours on pre-leisure days are not driven by medical factors. A simple back-of-the envelope calculation can shed some light on the potential gains that could result from reducing this practice. Being born during the normal shift on pre-leisure days increases the probability of C-section by 1.4 percentage points.[42] Given that the overall C-section rate in Finland from 1990 to 2014 is 16.45%, removing these excess cesareans that occur during pre-leisure days would lower the

C-section rate by 1.86%.[43] In 2014, there were 9,534 C-sections in total (out of 57,323 deliveries). This implies that, in absence of this practice, around 180 less C-sections would have been performed. Taking into account that the cost of a C-section without complications is 1,815 euros higher than that of a vaginal delivery,[44] eliminating this variation would result in about 326,700 euros of savings in just one year.

We provide this evidence in the context of Finland, one of the countries with the lowest C-section rate in the world (OECD, 2017). We would expect this variation to provide an even stronger source of identification in other countries with higher rates of medical interventionism during childbirth. Thus, this paper hopes to provide a solid base upon which future research on the effects of avoidable cesarean sections can be built.

---

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

[2]Understanding and quantifying the potential contribution of C-sections to the development of these diseases is not limited to medical practice and health policy. Chronic health conditions cause an immense financial burden to households and public health care financing. The total cost of asthma in the working age population was estimated to be $24.7 billion during 1999-2002 in Europe (Global Asthma Network, 2018). The two other atopic diseases we investigate imply high costs: atopic dermatitis has been estimated to cost at least $5.3 billion (in 2015 USD) in the US (Drucker et al., 2017). The estimated annual cost of allergic rhinitis is in the range of $2–5 billion (in 2003 USD) (Reed, Lee and McCrory, 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 in lifetime medical costs in the US (Finkelstein, Graham and Malhotra, 2014).

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

[4]To our knowledge, the only paper looking at longer-term effects is by Jachetta (2015). This paper explores the relationship between cesarean deliveries and hospitalizations using regional variation in medical malpractice insurance premia in the US as an instrument for C-sections.

However, the instrument used in that paper does not necessarily allow for credible causal inference, since the author finds that higher premia also predict delayed prenatal care, lower birth weight, and reduced gestational age.

[5]A few papers have also examined the effects of cesarean sections on mothers. Halla et al. (2020) study the effects of C-sections on fertility and maternal labor supply. Tonei (2019) studies the impact on mental health for mothers with breech babies who undergo a C-section. Our findings on children health complement these maternal results and contribute to obtaining a more complete picture of the effect of cesarean sections.

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

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

[8]The head of the department in each hospital has large discretionary power to assign physicians to their working schedules. However, physicians may often independently change their work shifts between themselves. Physicians working in public hospitals are generally not unilaterally able to choose their working schedules. Even though the statutes that govern on-call arrangements have changed in recent years, during most years covered in our data, small hospitals with less than 1000 annual births could autonomously decide their on-call arrangements. In certain hospitals, physicians were allowed to be at home while on duty, if they could arrive to the hospital within 30 minutes from home.

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

[10]Home births are also included in the registry but are extremely rare in Finland. From 1996 to 2013, there were 1,053,802 births in total, and only 197 were planned home deliveries (Ovaskainen et al., 2019). We focus on births that took place in hospitals.

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

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

[13]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.

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

[15]The vector of covariates includes parity, the gender of the baby, mother's marital status, nationality, socioeconomic status, age, and smoking status. In addition, we include pregnancy and delivery related indicators that include in-vitro fertilization, gestational weeks, high (above 75th percentile) and low (below 25th percentile) number of visits to prenatal clinic, induced labor, prostaglandin pre-induction, epidural use, and laughing gas anesthesia.

[16]Figure A1 in the Online Appendix shows that mothers and babies who undergo a C-section are very different from those mothers and babies who undergo a vaginal delivery.

[17]Working days that precede a leisure day include Fridays and days preceding public holidays. Table A4 in the Online Appendix documents all public holidays in Finland. Friday is not considered a working day that precedes a leisure day if it is a holiday.

[18]For our identification strategy we exploit only the variation presented in Figure 1b, that is, the excess C-sections performed during normal working hours on pre-leisure days compared to other working days. Figure 1a shows, however, that within any working day there is also substantial variation in the C-section rate by the time of birth. In particular, more cesareans are performed during normal working hours than during the rest of the day, and we observe the lowest probability of cesarean delivery during the early morning (from 1 am to 7 am). This suggests that physicians try to concentrate most obstetric surgeries during their regular working shift. We do not rely on this variation alone for identification.

[19]Figure A2 in the Online Appendix 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.

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

[21]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-medical mothers. 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.

[22]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, high (above 75th percentile) and low (below 25th percentile) number of visits to prenatal clinic, induced labor, prostaglandin pre-induction, epidural use, and laughing gas anesthesia.

[23]In Figure A3 in the Online Appendix, we explore whether the strength of the first stage varies by type of hospital. We do not find evidence that this variation is driven by any observable characteristic at the hospital level. The first-stage is very similar for hospitals of different sizes, for university and

non-university hospitals, for hospitals with low and high C-section rate, and for hospitals in high and low populated locations.

[24]While, ideally, we would like to construct our instrument based on the time of admission to the hospital, this information is not available in our data. One potential concern with using time of birth instead of time of admission is that time of birth might be affected by the type of delivery, as C-sections shorten labor. This could lead to compositional changes in the pool of mothers giving birth during the normal shift on different types of days. However, the comparison of a vast list of maternal characteristics in Figure 2 shows that any such compositional change does not appear to be quantitatively important in our context, as we do not find any significant difference among mothers giving birth during the normal shift on pre-leisure days compared to other working days. Moreover, Costa-Ramón et al. (2018) find that, in a similar context, using an instrument based on the time of birth instead of the time of admission does not affect the results in practice.

[25]It is important to note, however, that those numbers represent the current situation in a sample of hospitals and may not necessarily be an accurate description of the situation in previous years.

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

[27]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. The full list of controls is as follows: gender of the child, maternal age, maternal smoking status, maternal weight, and indicators for high/low number of prenatal visits, breech position, hospital visits during pregnancy due to eclampsia, hypertension and placenta previa, abnormal glucose levels, gestational weeks, preterm pregnancy, induced birth, prostaglandin pre-induction, epidural, laughing gas anesthesia, and diagnoses of prolonged and obstructed labor.

[28]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 only second children have C-sections in our sample.

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

[30]We perform this adjustment for each estimation method including both neonatal and long-run outcomes (all twelve outcome variables explored in Tables 2 and 3) using the rwolf Stata package (Clarke, 2016).

[31]The OLS estimation is ran in a sample that only excludes planned C-sections and births for which we do not observe parity. The specification includes the full set of controls and fixed effects described in Equation (1), as well as controls for birth order.

[32]We reviewed all the papers cited in the following meta-analyses: Bager, Wohlfahrt and Westergaard (2008); Cardwell et al. (2008); Darabi et al. (2019); Keag, Norman and Stock (2018); Li, Zhou and Liu (2013); Thavagnanam et al. (2008). Table A9 provides a full list of all papers reviewed for each outcome.

[33]In Figure A5 in the Online Appendix we show the number of papers that report the type of C-sections included. Out of 85 papers, 58 include all types of C-sections or do not report which group they include.

[34]The additional covariates included in our set of controls but not in most previous papers are: mother marital status, high or low number of visits to prenatal clinic during pregnancy, in-vitro fertilization, epidural use, laughing gas anesthesia, induced labor, and prostaglandin pre-induction.

[35]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).

[36]Shoulder dystocia is an obstetric emergency that is often unpredictable and is one of the most litigated causes in obstetrics (Politi et al., 2010). This occurs when the baby's head passes through the birth canal and their shoulders become stuck during labor. If this complication arises, effective and timely clinical management is essential to ensure the well-being of the newborn (Kwek and Yeo, 2006). Brachial plexus injury is one of the most important fetal complications of shoulder dystocia (Politi et al., 2010) and consists of an injury to the brachial plexus, that is, the network of nerves that conducts signals from the spinal cord to the shoulder, arm and hand.

[37]ICD-10 codes O85-O92.

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

[39]This is despite the fact that 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.

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

[41]We exploit the ICD-10 category: Injury, poisoning and certain other consequences of external causes, ICD-10 Codes: S00-T98.

[42]This is the first stage coefficient in column 2 in Table 1.

[43]We divide the first stage by the C-section rate and multiply the result by the percentage of days that were pre-leisure days in our sample (21.8%).

[44]These prices are based on the diagnosis-related group (DRG) pricing used by hospitals for internal pricing policies. We take the most recent price list in the largest hospital in Finland (Helsinki University Hospital). In this hospital, the DRG product price for a C-section (without complications) is 4,020 euros and for a vaginal delivery (without complications) 2,205 euros. The information can be found here (in Finnish, accessed on 14/05/2020): https://www.hus.fi/hus-tietoa/talous/Hinnoittelu/Documents/HUS%20Palveluhinnasto%202020,%20tuote-%20ja%20suoriteperusteiset%20hinnat%20(osat%201%20ja%202).pdf.

References

Aizer, Anna, Laura Stroud, and Stephen Buka. 2016. "Maternal Stress and Child Outcomes: Evidence from Siblings." *Journal of Human Resources*, 51(3): 523–555.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*, 56(4):1360–1446.

Almond, Douglas, Lena, Edlund and Mårten Palme. 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., S. Cnattingius, P. Lichtenstein, and C. Lundholm. 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.

Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics:An Empiricist's Companion*. Princeton University Press.

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

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

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

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

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

Black, Sandra E, Sanni Breining, David N Figlio, Jonathan Guryan, Krzysztof Karbownik, Helena Skyt Nielsen, Jeffrey Roth, and Marianne Simonsen. 2017. "Sibling Spillovers." National Bureau of Economic Research Working Paper 23062.

Blustein, Jan, and Jianmeng Liu. 2015. "Time to consider the risks of caesarean delivery for long term child health." BMJ, 350.

Boerma, Ties, Carine Ronsmans, Dessalegn Y. Melesse, Aluisio J.D. Barros, Fernando C. Barros, Liang Juan, Ann Beth Moller, Lale Say, Ahmad Reza Hosseinpoor, Mu Yi, Dácio de Lyra Rabello Neto, and Marleen Temmerman. 2018. "Global epidemiology of use of and disparities in caesarean sections." *The Lancet*, 392(10155): 1341–1348.

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

Card, David, Alessandra Fenizia, and David Silver. 2019. "The Impacts of Hospital Delivery Practices on Infant Health." *NBER Working Paper, 25986.*

Cardwell, C. R., L. C. Stene, G. Joner, O. Cinek, J. Svensson, M. J. Goldacre, R. C. Parslow, P. Pozzilli, G. Brigis, D. Stoyanov, B. Urbonaité, S. Šipetić, E. Schober, C. Ionescu-Tirgoviste, G. Devoti, C. E. de Beaufort, K. Buschard, and C. C. Patterson. 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, Richard C., Jishnu Das, and Michael Lokshin. 2012. "A practical comparison of the bivariate probit and linear IV estimators." *Economics Letters*, 117(3): 762 – 766.

Collado, Maria Carmen, Samuli Rautava, Erika Isolauri, and Seppo Salminen. 2015. "Gut microbiota: a source of novel tools to reduce the risk of human disease?" *Pediatric Research*, 77: 182 – 188. Review.

Costa-Ramón, Ana María, Ana Rodríguez-González, Miquel Serra-Burriel, and Carlos Campillo-Artero. 2018. "It's about time: Cesarean sections and neonatal health." *Journal of Health Economics*, 59: 46 – 59.

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

Darabi, Behzad, Shoboo Rahmati, Mohammad Reza HafeziAhmadi, Gholamreza Badfar, and Milad Azami. 2019. "The association between caesarean section and childhood asthma: an updated systematic review and meta-analysis." *Allergy, Asthma & Clinical Immunology*, 15(1): 62.

Dominguez-Bello, Maria G, Kassandra M De Jesus-Laboy, Nan Shen, Laura M Cox, Amnon Amir, Antonio Gonzalez, Nicholas A Bokulich, Se Jin Song, Juana I Rivera-vina, Keimari Mendez, Rob Knight, and Jose C Clemente. 2016. "Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer." *Nature Medicine*, 22(3): 250–253.

Drucker, Aaron M., Annie R. Wang, Wen Qing Li, Erika Sevetson, Julie K. Block, and Abrar A. Qureshi. 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., D. A. Richardson, J. S. Sholl, and B. A. Johnson. 1984. "Cesarean section. Assessment of the convenience factor." *J Reprod Med*, 29(9): 670–676.

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

Finnish Medical Association. 2016. "Physicians in Finland, Statistics on Physicians and the Health Care System 2016." Finnish Medical Association Available at: https://www.laakariliitto.fi/site/assets/files/1268/ll16_tilasto2016_net1_170114.pdf (Accessed 14.5.2020).

Fraser, William, Robert H. Usher, Frances H. McLean, Carol Bossenberry, Mary Ellen Thomson, Michael S. Kramer, L. Paul Smith, and Hugh Power. 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*. Auckland, New Zealand:Global Asthma Network.

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

Halla, Martin, Harald Mayr, Gerald J. Pruckner, and Pilar Garcia-Gomez. 2020. "Cutting Fertility? The Effect of Cesarean Deliveries on Subsequent Fertility and Maternal Labor Supply." *Journal of Health Economics*, forthcoming.

Hannah, Mary E, Walter J Hannah, Sheila A Hewson, Ellen D Hodnett, Saroj Saigal, Andrew R Willan, and Term Breech Trial Collaborative. 2000. "Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial." *The Lancet*, 356(9239): 1375–1383.

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

Hyde, Matthew J., Alison Mostyn, Neena Modi, and Paul R. Kemp. 2012. "The health implications of birth by Caesarean section." *Biological Reviews*, 87(1): 229–243.

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

Jachetta, Christine. 2015. "Cesarean Sections and Later Health Outcomes.". Unpublished, Available at http://www.christinejachetta.com.

Jensen, Vibeke Myrup, and Miriam Wüst. 2015. "Can Caesarean section improve child and maternal health? The case of breech babies." *Journal of Health Economics*, 39: 289 – 302.

Johnson, Erin M., and M. Marit Rehavi. 2016. "Physicians Treating Physicians: In- formation and Incentives in Childbirth." *American Economic Journal: Economic Policy*, 8(1): 115–41.

Keag, Oonagh E., Jane E. Norman, and Sarah J. Stock. 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.

Kuhle, Stefan, Sara F L Kirk, Arto Ohinmaa, and Paul J Veugelers. 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.

Kwek, Kenneth, and George S. H. Yeo. 2006. "Shoulder dystocia and injuries: prevention and management." *Current Opinion in Obstetrics and Gynecology*, 18(2).

Li, HT, YB Zhou, and JM Liu. 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., L. Baur, and R. Uauy. 2004. "Obesity in children and young people: a crisis in public health." *Obesity Reviews*, 5(s1): 4–85.

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

Mishanina, Ekaterina, Ewelina Rogozinska, Tej Thatthi, Rehan Uddin-Khan, Khalid S. Khan, and Catherine Meads. 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, Hanna. 2017. Cesarean section for high-risk births: short term consequences for breech births." https://www.dropbox.com/s/xto1qlsf1qtxf7d/C-section_Muhlrad.pdf. Accessed: 26-12-2017.

Neu, Josef, and Jona Rushing. 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." National Institute for Health and Care Excellence NICE Clinical Guideline CG190. https://www.nice.org.uk/guidance/cg132/chapter/1-guidance. Accessed: 06-12-2017.

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

Nielsen, Helena Skyt, Nina Smith, and Aycan Çelikaksoy. 2009. "The effect of marriage on education of immigrants: Evidence from a policy reform restricting marriage migration." *Scandinavian Journal of Economics*, 111(3): 457–486.

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

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

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

Ovaskainen, Katja, Riitta Ojala, Kati Tihtonen, Mika Gissler, Tiina Luukkaala, and Outi Tammela. 2019. "Planned home deliveries in Finland, 1996-2013." *Journal of Perinatology*, 39(2): 220–228.

Peters, Lilian L., Charlene Thornton, Ank de Jonge, Ali Khashan, Mark Tracy, Soo Downe, Esther I. Feijen-de Jong, and Hannah G. Dahlen. 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*, 1–11.

Politi, Salvatore, Laura D'emidio, Pietro Cignini, Maurizio Giorlandino, and Claudio Giorlandino. 2010. "Shoulder dystocia: an Evidence-Based approach." *Journal of Prenatal Medicine*, 4(3): 35–42.

Prior, Emily, Shalini Santhakumaran, Chris Gale, Lara H Philipps, Neena Modi, and Matthew J Hyde. 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, Shelby D, Todd A Lee, and Douglas C McCrory. 2004. "The Economic Burden of Allergic Rhinitis." *PharmacoEconomics*, 22(6): 345–361.

Saccone, Gabriele, and Vincenzo Berghella. 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.
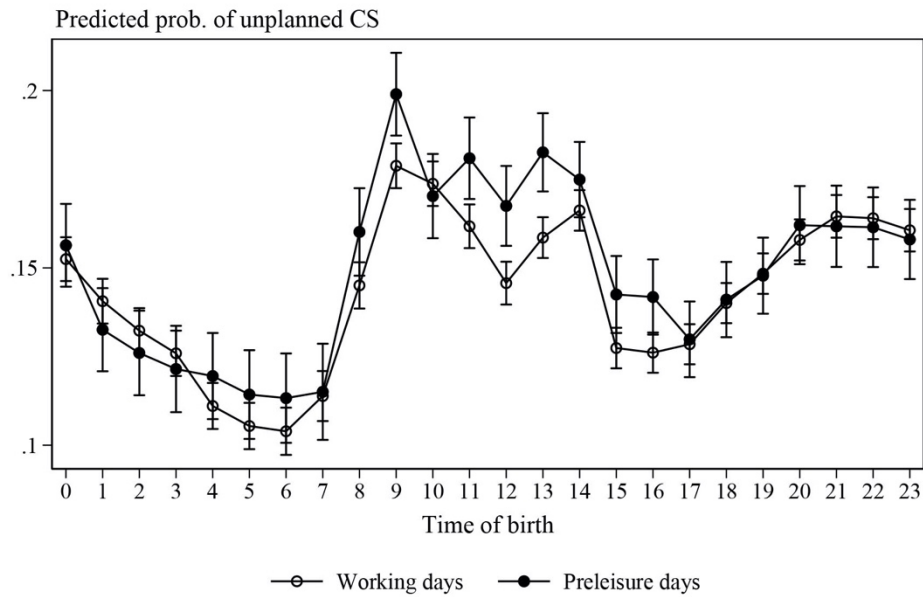
Scudellari, Megan. 2017. "News Feature: Cleaning up the hygiene hypothesis." *Proceedings of the National Academy of Sciences*, 114(7): 1433–1436.

Stock, James, and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models*. , ed. Donald W.K. Andrews, 80–108. New York: Cambridge University Press.
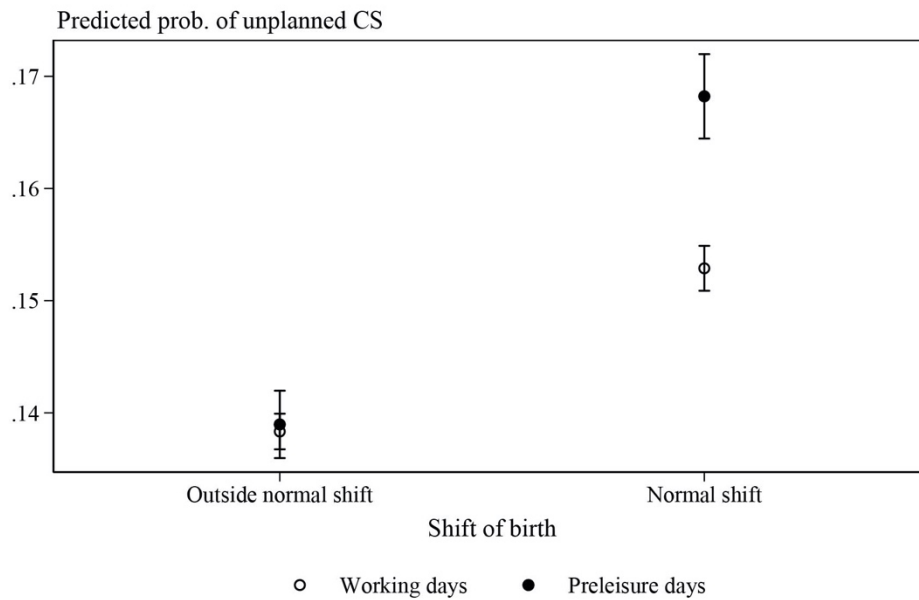
Strachan, D P. 1989. "Hay fever, hygiene, and household size." BMJ *(Clinical research ed.),* 299(6710): 1259–60.

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

Tao, Betty, Massimo Pietropaolo, Mark Atkinson, Desmond Schatz, and David Taylor. 2010. "Estimating the cost of type 1 diabetes in the U.S.: A propensity score matching method." *PLoS ONE*, 5(7).

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

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

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

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

Figures (to be placed in the article)

Predicted prob. of unplanned CS

Time of birth

—○— Working days —●— Preleisure days

(a) By time of birth

Predicted prob. of unplanned CS

Shift of birth

○ Working days ● Preleisure days

(b) By shift of birth

Figure 1
Predicted probability of unplanned C-section
Notes: Figure (a) presents the predicted probability of unplanned C-section by hour and type of day. Figure (b) shows the predicted probability of unplanned C-section by shift and type of day. Both figures adjust for hospital, month, and year of birth fixed effects. Pre-leisure days include working days that precede a Finnish public holiday or a weekend, while working days include the rest of working days. Sample is restricted to singleton first births which are either unscheduled C-sections or vaginal births.
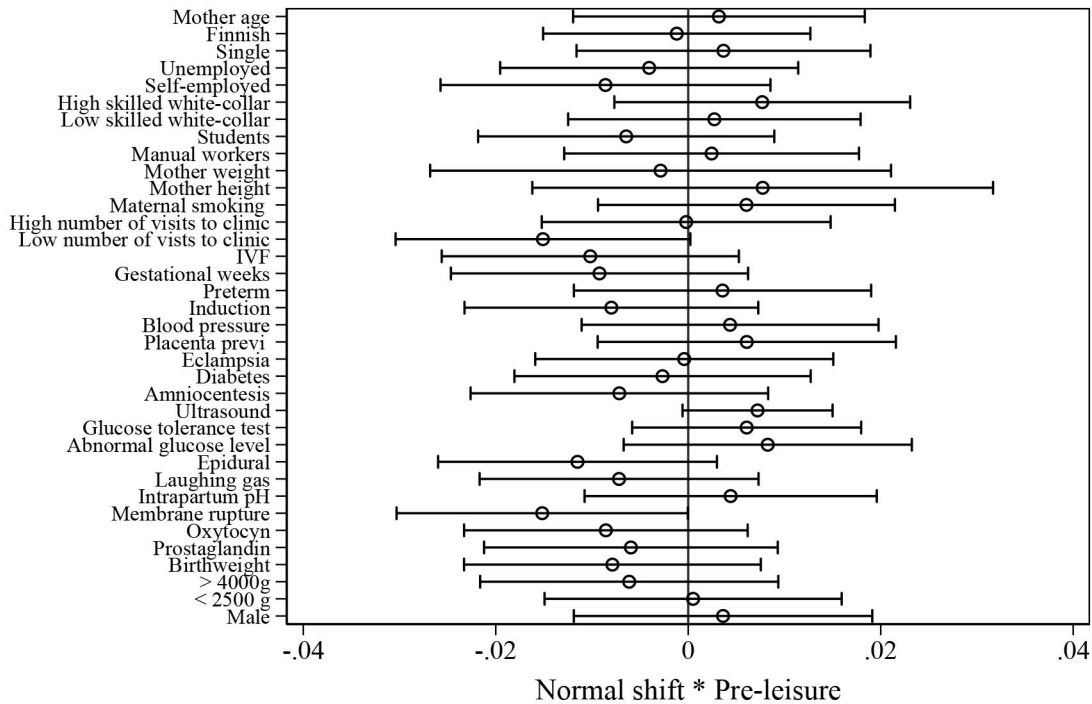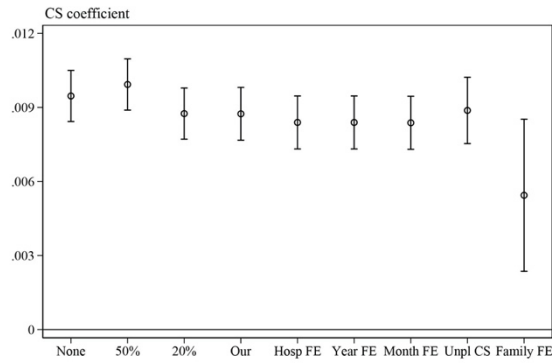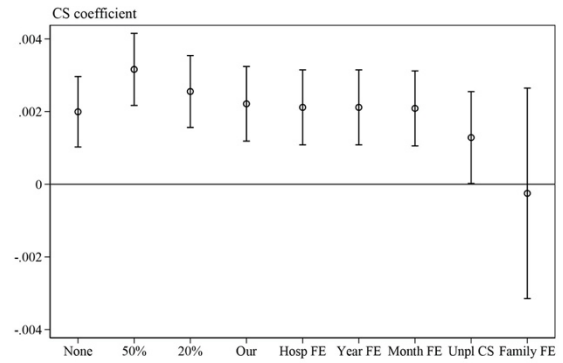
Figure 2

Instrument and baseline characteristics

Notes: The figure represents the coefficients and 95% CI from separate regressions of each predetermined variable (standardized to have mean 0 and standard deviation 1) on the instrument (Normal shift * Pre-leisure), controlling for normal shift time, pre-leisure day, and hospital, month, and year of birth fixed effects. Sample is restricted to single births, unscheduled C-sections and vaginal births that take place on working days.
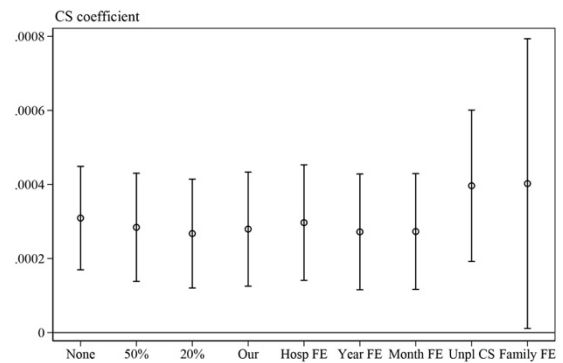
(a) Asthma

(b) Atopy
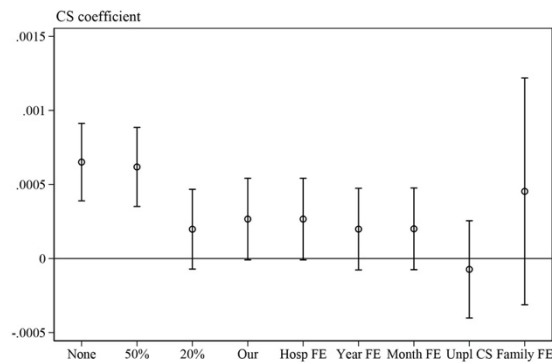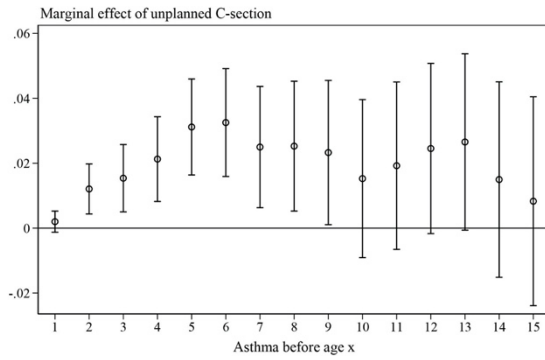

(c) Type 1 Diabetes

(d) Obesity

Figure 3

OLS estimation – By group of controls

Notes: These figures plot the cesarean section coefficients and their 95% CI from the OLS estimation of the association between C-section and the probability of being diagnosed with asthma, atopy, type 1 diabetes, or obesity before age 5. In each figure, each coefficient corresponds to a different specification with varying sets of controls which are sequentially added: no controls in the first; the list of controls included in at least 50% of medical studies in the second; controls included in at least 20% of them in the third (more information in Table A9); the rest of maternal, pregnancy, and delivery characteristics included in our usual group of controls (described in Section III.B) in the fifth; hospital FE in the sixth; year FE in the seventh, month FE in the eighth; excluding planned C-sections in the ninth, and finally including family FE in the last column.

Marginal effect of unplanned C-section

(a) Asthma

Marginal effect of unplanned C-section

(b) Atopy

Marginal effect of unplanned C-section

(c) Type 1 Diabetes
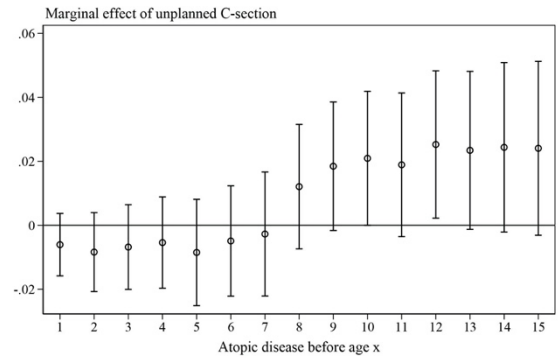
Marginal effect of unplanned C-section
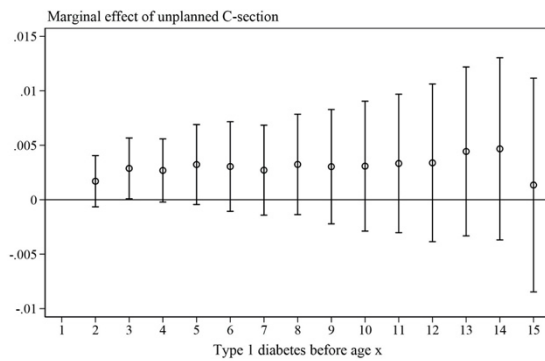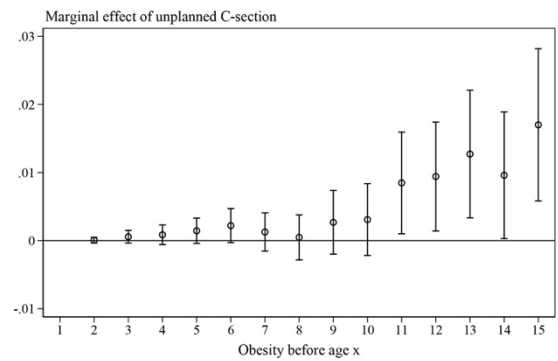
(d) Obesity

Figure 4

Bivariate probit estimation – Child diagnoses by age

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 III.B.1.
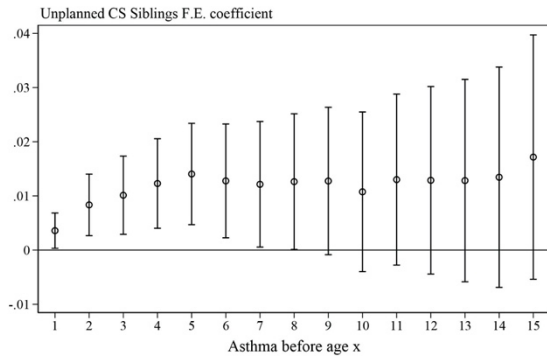
(a) Asthma
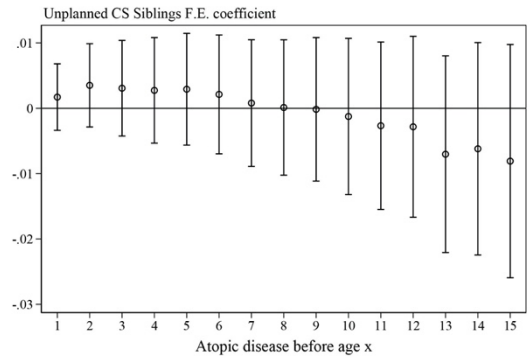
(b) Atopy
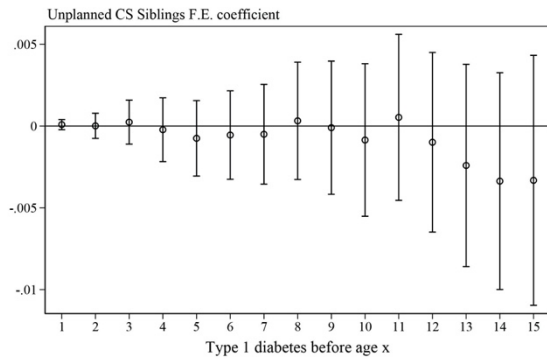
(c) Type 1 Diabetes

(d) Obesity

Figure 5

Diff-in-diff analysis – Child diagnoses by age

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 III.B.2.

Tables (to be placed in the article)

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
Mean of Y	0.145	0.145
First-stage F Statistic	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 pre-induction, epidural or laughing gas anesthesia. Robust standard errors in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Table 2
Neonatal outcomes

	(1) Low Apgar 1	(2) ICU	(3) Assisted ventilation	(4) Neonatal mortality
Ordinary Least Squares	0.068*** (0.001) [0.002]	0.118*** (0.001) [0.002]	0.027*** (0.001) [0.002]	0.002*** (0.000) [0.002]
Mean of Y	0.049	0.087	0.009	0.001
Observations	1119467	1120932	1120932	1119842
Two-Stage Least Squares	-0.018 (0.140) [1.000]	-0.088 (0.170) [0.999]	-0.006 (0.061) [1.000]	0.006 (0.023) [1.000]
Mean of Y	0.066	0.106	0.012	0.002
Observations	392017	392560	392560	392173
Bivariate probit	0.104*** (0.008) [0.023]	0.163*** (0.009) [0.001]	0.017*** (0.005) [0.297]	-0.001 (0.005) [0.927]
Mean of Y	0.066	0.106	0.012	0.002
Observations	392017	392560	392560	392173
Difference-in-differences	0.053*** (0.007) [0.000]	0.111*** (0.007) [0.000]	0.036*** (0.004) [0.000]	0.001 (0.002) [0.919]
Mean of Y	0.038	0.070	0.006	0.001
Observations	644551	645292	645292	644746
First-stage F Statistic	24.996	25.216	25.216	26.007

Notes: This table shows the estimates of the marginal effect of an unplanned CS on different neonatal health indicators by OLS, 2SLS, bivariate probit, and differences-in-differences estimation (see Equations (1), (2), (3), and (6)). Specifications as detailed in Section III.B, with the full set of fixed effects and controls. The sample for bivariate probit and 2SLS includes only singleton first births; for diff-in-diff, both first and second singleton births from families where the first child is born by vaginal delivery (and the second is born by vaginal delivery or unplanned C-section). The OLS estimation includes both first and higher order singleton births. Robust standard errors (in parentheses) in panels 1-3, and standard errors clustered at the family level in the differences-in-differences panel. First-stage F statistic from 2SLS and bivariate probit specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ represent standard p-values. Romano-Wolf multiple hypothesis adjusted p-values in square brackets.

Table 3
Child diagnoses by age

	Type 1 diabetes		Asthma		Obesity		Atopy	
By age:	5	10	5	10	5	10	5	10
Ordinary Least Squares	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)
	[0.585]	[0.361]	[0.002]	[0.002]	[0.008]	[0.002]	[0.042]	[0.006]
Mean of Y	0.003	0.006	0.045	0.071	0.001	0.004	0.044	0.061
Observations	807035	556009	807035	556009	807035	556009	807035	556009
Two-Stage Least Squares	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)
	[0.072]	[0.786]	[0.996]	[0.990]	[1.000]	[1.000]	[1.000]	[0.990]
Mean of Y	0.003	0.006	0.040	0.070	0.001	0.004	0.041	0.058
Observations	296998	217768	296998	217768	296998	217768	296998	217768
Bivariate probit	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)
	[0.504]	[0.727]	[0.005]	[0.727]	[0.572]	[0.727]	[0.727]	[0.413]
Mean of Y	0.003	0.006	0.040	0.070	0.001	0.004	0.041	0.058
Observations	296998	217768	296998	217768	296998	217768	296998	217768
Difference-in-Differences	-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)
	[0.919]	[0.933]	[0.005]	[0.287]	[0.532]	[0.933]	[0.919]	[0.933]
Mean of Y	0.003	0.006	0.045	0.070	0.001	0.004	0.044	0.060
Observations	510075	366885	510075	366885	510075	366885	510075	366885
First-stage F Statistic	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 (marginal effects), and differences-in-differences estimation (see Equations (1) (2), (3), and (6)). Specifications as detailed in Section III.B, with the full set of fixed effects and controls. The sample for bivariate probit and 2SLS includes only singleton first births; for diff-in-diff, both first and second singleton births from families where the first child is born by vaginal delivery (and the second is born by vaginal delivery or unplanned C-section). The OLS estimation includes both first and higher order singleton births. Robust standard errors (in parentheses) in panels 1-3, and standard errors clustered at the family level in the differences-in-differences panel. First-stage F statistic from 2SLS specifications. * p < 0.1, ** p < 0.05, *** p < 0.01 represent standard p-values. Romano-Wolf multiple hypothesis adjusted p-values in square brackets.

Table 4
Exclusion restriction validity checks

	Planned birth	Shoulder dystocia	Brachial plexus injury	Length of stay	Maternal complications
Normal shift	0.130***	0.000	-0.000**	0.031*	0.001*
	(0.001)	(0.000)	(0.000)	(0.018)	(0.000)
Preleisure day	0.013***	0.000	-0.000	0.022	0.000
	(0.001)	(0.000)	(0.000)	(0.024)	(0.000)
Normal shift*Preleisure	0.001	-0.000*	-0.000	0.005	-0.001*
	(0.002)	(0.000)	(0.000)	(0.037)	(0.001)
Observations	1095943	1095943	1095943	53633	392560
Adjusted R^2	0.083	0.003	0.002	0.225	0.009
Controls	YES	YES	YES	YES	YES
Mean of Y	0.236	0.001	0.002	5.226	0.010
First-stage F Statistic	0.335	2.904	0.113	0.019	2.936

Notes: This table shows the results of regressing different proxy measures of quality of care on the instrument (Normal shift*Preleisure), controlling for normal shift, pre-leisure day, hospital, year, and month of birth fixed effects and the full set of controls as described in Equation (2). The outcome in column 1 is a dummy equal to 1 if the birth was planned (planned C-section or induction) and 0 otherwise. In column 2, the dependent variable is a dummy for shoulder dystocia during childbirth, and in column 3, a dummy for brachial plexus injury. In column 4, the dependent variable is mother's length of stay (in days) and the sample is restricted to mothers who delivered by C-section. Finally, in column 5, the outcome is a dummy equal to 1 if the mother is diagnosed with any complication related to the puerperium within one week from the delivery (ICD-10 Codes: O85-O92). Here the sample is restricted to single births, unscheduled C-sections, and vaginal births that take place on working days. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$,*** $p < 0.01$

Table 5
Validity checks

	Birth weight	Asthma at age 5 for sample	
		Thursdays vs Fridays	Excluding inductions
Bivariate probit	-	0.023	0.036***
	-	(0.015)	(0.010)
Mean of Y	-	0.040	0.039
Observations	-	117826	246933
Difference-in-differences	-5.416	-	0.017**
	(7.617)	-	(0.007)
Mean of Y	3566.117	-	0.044
Observations	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 the differences-in-differences (bottom) estimation of the impact of unplanned CS on the probability of asthma diagnosis by age 5 restricting the sample to births taking place on Thursdays or Fridays (col. 2) or to non-induced births (col. 3). Specifications as detailed in Sections III.B.1 and III.B.2, with the full set of fixed effects and controls. Robust standard errors (in parentheses) for bivariate probit results, and standard errors clustered at the family level in the differences-in-differences panel. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6
Validity of differences-in-differences

	Neonatal health				Diagnosis by age 5			
	Low Apgar	ICU	Assisted Ventilation	Neonatal mortality	Type 1 diabetes	Asthma	Obesity	Atopy
Instrumented	0.060***	0.020**	0.001	-0.001	-0.001	-0.004	-0.001	-0.006
	(0.009)	(0.009)	(0.003)	(0.002)	(0.002)	(0.008)	(0.001)	(0.009)
Mean of Y	0.028	0.061	0.005	0.001	0.003	0.044	0.001	0.044
Observations	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)
Mean of Y	0.035	0.062	0.005	0.001	0.003	0.044	0.001	0.044
Observations	608688	609368	609368	608909	482536	482536	482536	482536

Notes: This table shows the results from sibling fixed effect models, following the specification in Equation (6), for two different samples of children: in the top panel, for a sample of sibling pairs where the first child was born by eutocic birth, and the second child is born either by eutocic or instrumented vaginal birth; in the bottom panel, for vaginally delivered sibling pairs where the first child did not have a high-risk pregnancy and the second child had a low- or high-risk pregnancy. The top panel coefficient represents the change in the health gap between siblings in families where the second child was born by instrumented vaginal delivery, while the bottom panel coefficient represents the same for families where the second child had a high-risk pregnancy. All specifications include family, hospital, year, and month of birth fixed effects and the controls described in section III.B.2. Standard errors are clustered at the family level. * p < 0.1, ** p < 0.05, *** p < 0.01