Universitat Pompeu Fabra Barcelona

CRES CENTRE DE RECERCA EN ECONOMIA I SALUT

IT'S ABOUT TIME: CESAREAN SECTIONS AND NEONATAL HEALTH†

Ana María Costa-Ramón ‡
Universitat Pompeu Fabra & CRES

Ana Rodríguez-González
Universitat Pompeu Fabra

Miquel Serra-Burriel
Universitat de Barcelona & CRES

Carlos Campillo-Artero
Servei de Salut de les Illes Balears & CRES

September 17

CRES-UPF Working Paper #201709-94

It's About Time: Cesarean Sections and Neonatal Health†

Ana María Costa-Ramón ‡
Universitat Pompeu Fabra & CRES

Ana Rodríguez-González
Universitat Pompeu Fabra

Miquel Serra-Burriel
Universitat de Barcelona & CRES

Carlos Campillo-Artero
Servei de Salut de les Illes Balears & CRES

Abstract

Cesarean sections have been associated in the literature with poorer newborn health, particularly with a higher incidence of respiratory morbidity. Most studies suffer, however, from potential omitted variable bias, as they are based on simple comparisons of mothers who give birth vaginally and those who give birth by cesarean section. We try to overcome this limitation and provide credible causal evidence by using variation in the probability of having a c-section that is arguably unrelated to maternal and fetal characteristics: variation by time of day. Previous literature documents that, while nature distributes births and associated problems uniformly, time-dependent variables related to physicians' demand for leisure are significant predictors of unplanned c-sections. Using a sample of public hospitals in Spain, we show that the rate of c-sections is higher during the early hours of the night compared to the rest of the day, while mothers giving birth at the different times are similar in observable characteristics. This exogenous variation provides us with a new instrument for type of birth: time of delivery. Our results suggest that non-medically indicated c-sections have a negative and significant impact on newborn health, as measured by Apgar scores and the pH level of the umbilical cord.

Keywords: Cesarean section, c-section, neonatal health, time variation, instrumental variables.
JEL Codes: I10, I12.

† We are grateful to Libertad González, Guillem López-Casasnovas, Cristina Bellés-Obrero, Andrés Calvo, Rosa Ferrer, Christian Fons-Rosen, Borja García Lorenzo, Albrecht Glitz, Sergi Jiménez-Martín, Gianmarco León, Vicente Ortún, Alexandrina Stoyanova, Alessandro Tarozzi and Ana Tur-Prats. We also thank participants in the UPF LPD Seminar, VI EvaluAES Workshop, 31st ESPE Conference and 12th iHEA Congress.

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

1 Introduction

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

This paper aims to help fill this research gap by providing new evidence of a causal link between unplanned cesarean sections and newborn health outcomes. Understanding the impact of c-sections on neonatal health is of relevance, as fetal and neonatal outcomes have been shown to be determinants not only of future health, but also of other later life outcomes, such as test scores, educational attainment, and income (Almond and Currie, 2011). In particular, we look at the impact of c-sections on Apgar scores and on the pH of the umbilical cord, both widely used measures of newborn well-being. Apgar scores have been found to be predictive of health, cognitive ability, and behavioral problems of children at age three (Almond et al., 2005), and of reading and math test scores in grades 3-8 (Figlio et al., 2014). The pH of the umbilical cord, in turn, is stressed in the medical literature as an important outcome measure, due to its strong association with neonatal mortality and morbidity (Malin et al., 2010).

In order to show the existence of a causal link between unscheduled c-sections and health, we use exogenous variation in the probability of having a c-section at different times of day. Indeed, although nature distributes births and associated problems uniformly, some studies have demonstrated that time-dependent variables related to physicians' demand for leisure are significant predictors of unplanned c-sections (Brown, 1996). Using a sample of birth registries in public hospitals in Spain, we first document that, in this context, emergency c-sections are more likely to be performed in the early hours of the night (from 11 pm to 4 am). We discuss how the structure of medical shifts and the higher opportunity cost in terms of time that vaginal deliveries imply might explain physicians' incentives to perform more c-sections during this time of day. We then show that mothers giving birth at different times of day are observationally similar, also in terms of pregnancy and labor characteristics that might predict a medically-indicated c-section. The results thus suggest that the excess number of c-sections observed at the early night are due to non-medical reasons. We consequently adopt an instrumental variable approach, using time of birth as an instrument for the mode of delivery. This allows us to interpret our estimates as causal and to focus on avoidable c-sections, as medically-indicated cesareans will be performed independently of the time of birth. We discuss the necessary assumptions and their plausibility. Our results suggest that non-medically indicated c-sections lead to a significant worsening of Apgar scores of approximately one standard deviation, and an increased probability of having the pH of the umbilical cord below normal levels. Our findings are robust to a number of robustness checks.

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

To the best of our knowledge, the only paper that endeavors to identify the causal impact of cesareans on later infant health is Jachetta (2015). The author uses variation in medical malpractice premia at the Metropolitan Statistical Area (MSA) level in the US as an instrument for the rate of risk-adjusted cesarean sections and finds that higher rates lead to an increase in the rate of total hospitalizations and of hospitalizations that present asthma. Although the author identifies several potential threats to the validity of the instrument, the paper is a first step towards providing credible estimates of the causal link between c-sections and health outcomes. We advance the existing knowledge by using a new instrument that allows us to isolate the causal impact of non-medically indicated c-sections on newborn health. In particular, our setting allows us to focus on mothers that give birth in the same hospital and have similar observable characteristics, differing only in the time of delivery. Moreover, we are able to provide evidence that time-variation in the quality of care is not driving our results. Finally, because we measure the impact on health at birth, we are able to establish a direct connection between c-sections and health outcomes.

Second, our work is also related to the literature that documents or uses time variation in the probability of having a c-section. Brown (1996) was one of the first to show that the probability of unplanned c-sections is non-uniformly distributed across time. Using data from military hospitals in the US, the author finds that cesarean sections were less likely to occur during the weekend and more likely from 6 pm to 12 am. He interprets these results as evidence that non-clinical variables, in particular physicians' demand for leisure, also play a role in doctors' decision-making. In our setting, we find that the probability of unplanned c-sections is higher during the early hours of the night. It is during this time that doctors appear to have a higher incentive to perform a c-section when facing ambiguous cases, as the opportunity cost in terms of time for a vaginal delivery is higher.

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

The structure of the rest of the paper is as follows. In the next section we provide background information on the choice of mode of delivery, on the institutional setting and physicians' shifts, and on why we would expect to find an adverse effect of c-sections on health outcomes. The third section introduces the data, describes the variation in the c-section rate across a 24-hour cycle and presents the empirical strategy. In section 4 we show and discuss our results. Section 5 presents some robustness checks and, finally, section 6 concludes.

2 Background

2.1 Choice of the mode of delivery

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

If there is no scheduled c-section, an attempt of vaginal delivery begins with the onset of labour or medical induction. If before or during labor the midwife or doctor detects evident health risks for the mother or the fetus, then a medically indicated emergency c-section will be performed. In some cases, however, whether or not a c-section is needed is not obvious, and thus the choice between a vaginal delivery or a c-section will depend on the subjective assessment of the doctor (Halla et al., 2016). As Shurtz (2013) points out, a c-section is a common procedure known to be sensitive to physician incentives. Several papers have found, for example, that financial fees can influence doctors' behavior (Grant, 2009). When fees are higher for a c-section than for a vaginal delivery, physicians have a greater incentive to perform a c-section. Other studies suggest that physicians perform more c-sections as a defensive strategy reflecting a fear of malpractice lawsuits (Baicker et al., 2006; Currie and MacLeod, 2008; Jachetta, 2015). Finally, physicians have more incentives to perform c-sections when the opportunity cost of time is higher, as vaginal deliveries take longer than c-sections and thus the latter can be seen as a time-saving device (Lefèvre, 2014). We focus here on this last type of incentive given that, by performing our analysis within hospital and exploiting variation across time of day, we abstract from variations in malpractice premia and financial fees.

2.2 Mechanisms: the impact of c-sections on newborn health

Cesarean sections have been associated with several adverse health outcomes for newborns. Hyde et al. (2012) provide an extensive review of such findings, concluding that although further research is needed, the available evidence suggests that "normal vaginal delivery is an important programming event with life-long health consequences." More specifically, the absence or modification of a vaginal delivery has been linked to several health alterations, which they classify as either short- or long-term.

The most relevant for our study, among the short-term outcomes, include the increased risk of impaired lung functioning and altered behavioral responses to stress. With regard to the former, one of the most common causes of respiratory distress among newborns is transient tachypnea or the presence of retained lung fluid. While in the amniotic sac, a baby's lungs are filled with amniotic fluid, but during labor the baby releases chemicals which, together with the pressure of the birth canal on the baby's chest, help expel the amniotic fluid from their lungs. This process does not occur when babies are born by cesarean section, such that the presence of fluid in their lungs after birth is more common. Moreover, catecholamines, one of the chemicals released by the fetus during labor, are also correlated with muscle tone and excitability. Otamiri et al. (1991) find that babies born by cesarean section responded worse to neurological tests a few days after birth. In our setting, we can proxy the impact of c-sections on these outcomes by looking at Apgar scores at minute 1 and 5 after birth, which capture appearance (skin color), pulse (heart rate), grimace (reflex irritability), activity (muscle tone), and respiration.

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

In addition to Apgar scores, we also study the impact of cesarean sections on the pH of the umbilical cord. The examination of the umbilical artery provides a measure of fetal distress and determines whether an infant has experienced an oxygen-depriving event. PH values below 7.20 reflect that the newborn suffered a moderate lack of oxygen and values under 7.15 suggest a severe lack of oxygen. Although the relationship between pH levels and Apgar scores is not one-to-one, they are positively correlated[1]. The medical literature recommendation is to consider pH levels together with Apgar scores in order to assess the well-being of the newborn (Hannah, 1989; Gao et al., 2009; Malin et al., 2010).

2.3 Institutional setting

2.3.1 Childbirth in Spanish public hospitals

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

Women may opt for private care, but most deliveries – 8 out of 10 births – take place under the public health system (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2015). In the year 2014, the c-section rate in the public health system was 22.1%, lower than the 25.4% rate of the whole sector, combining both public and private hospitals (*ibid.*). It is important to note that within the public system, obstetricians' wages are independent of the method of delivery used or the number of c-sections performed.

[1]Figure A1 in the appendix shows the distributions of umbilical cord pH for infants with Apgar scores 1 above and below 9 (first panel), and for infants with Apgar scores 5 above and below 9 (second panel). We observe that the distribution of pH levels for infants with Apgar scores below 9 is shifted to the left compared to that for babies with higher scores, with this being more salient for Apgar score 5.

2.3.2 Physicians' shifts

In our setting, the typical work shift for a doctor is from 8 am to 3 pm; night shifts are covered by doctors that are on duty and must stay in the hospital for 24 hours (from 8 am to 8 am next morning). When doctors are on duty, they provide assistance in (relatively uncommon) gynecological emergencies, occasionally monitor newborns' health, and are present in the labor room when decisions regarding a delivery are made, or if complications arise. Midwives, on the other hand, work 12-hour shifts (from 8 am to 8 pm).

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

3 Data and methods

3.1 Description of the data

Our data consists of all 6,163 birth records from four public hospitals in different Autonomous Regions in Spain during the years 2014-2016. The characteristics of the hospitals in our sample are comparable to that of the majority of public hospitals in Spain, in particular with regard to the volume of births attended per year (between 300 and 1500). In terms of c-section rates, three of the four hospitals are in the left tail of the distribution, while one is just at the mode, with a c-section rate around 21%. This comparison can be found in figure A2 in the appendix.

Each birth registry contains information on the mother's characteristics (age, nationality, education, marital status, etc.), on the pregnancy, on the type of birth (elective cesarean, emergency cesarean, eutocic delivery, etc.), on medical interventions during labor, on a series of medical indicators collected before, during, and after the delivery, on the newborn (birthweight, Apgar scores, etc.), and on the date and time of birth. Table A1 shows some summary statistics of the variables of interest. In our data, 5% of women delivered via a planned c-section, more than 11% via an emergency c-section, and 68% had an eutocic delivery, that is, a vaginal delivery without other interventions (i.e. spatula, forceps, or vacuum). Vaginal deliveries with such interventions represent around 15% of the sample. We restrict our sample to single births that are either eutocic or by unplanned c-section: our final sample consists of 4,886 observations.

3.2 Variation in the c-section rate by time of day

Figure 1 shows the c-section rate at different times of day for our sample of public hospitals in Spain. We can observe that the distribution of emergency c-sections by time of birth is not uniform. The proportion of women that deliver via an emergency c-section is higher in the early hours of the night (from 11 pm to 4 am), and much lower during the remaining hours of the night and the rest of the day. This pattern is not matched by either the total number of births or the number of vaginal births (see figure A3 in the appendix). More importantly, this variation is not driven by differences in maternal or pregnancy characteristics of the deliveries that take place at different times of day. In the next section, Table 2 confirms the balance of a very large set of mother and pregnancy characteristics between women delivering in the early hours of the night and during the rest of the day. As we will discuss in further detail, this allows us to use this exogenous variation as an instrument for mode of delivery.

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

This explanation is consistent with the time pattern that we observe in our data. Given the medical shift structure and the larger time-cost of surveillance implied by vaginal deliveries, doctors' incentives to perform c-sections in ambiguous cases may vary by time of day. In particular, we expect doctors to have a larger incentive to perform c-sections in the early hours of the night. By this time, on-duty doctors have already been working for more than 12 straight hours (see Figure A4 in the appendix). If they perform a c-section and do not have other mothers to care for, they can expect to rest for the remainder of their shift. Alternatively, if they do not perform a c-section, they will need to occasionally monitor the vaginal delivery throughout the night. Moreover, ongoing deliveries in the early hours of the night have a high probability of falling under the responsibility of the doctor on duty[2], as opposed to deliveries which begin later and are more likely to finish past the doctor's shift. These conditions would suggest that a higher share of deliveries with ambiguous indications end up as cesarean sections during the early hours of the night, as compared to the rest of the day.

[2] Average duration for the first stage of labor in vaginal deliveries among first-time mothers is around 8 hours (NICE, 2014).

Other alternative explanations are not compatible with this variation. For example, if either patient's or physician's fatigue increased the probability of c-sections, we would expect to see a higher emergency c-section rate during the late hours rather than the early hours of the night. We can also rule out that this is driven by an accumulation of births during these hours, as we do not observe the same time pattern for the number of births (see figure A3 in the Appendix). Finally, the early night spike in c-sections cannot be explained by selection of highly interventionist doctors at different times of day, as deliveries are not pre-assigned to a given obstetrician. We also show evidence that this is not the case in Figure A5 in the appendix.

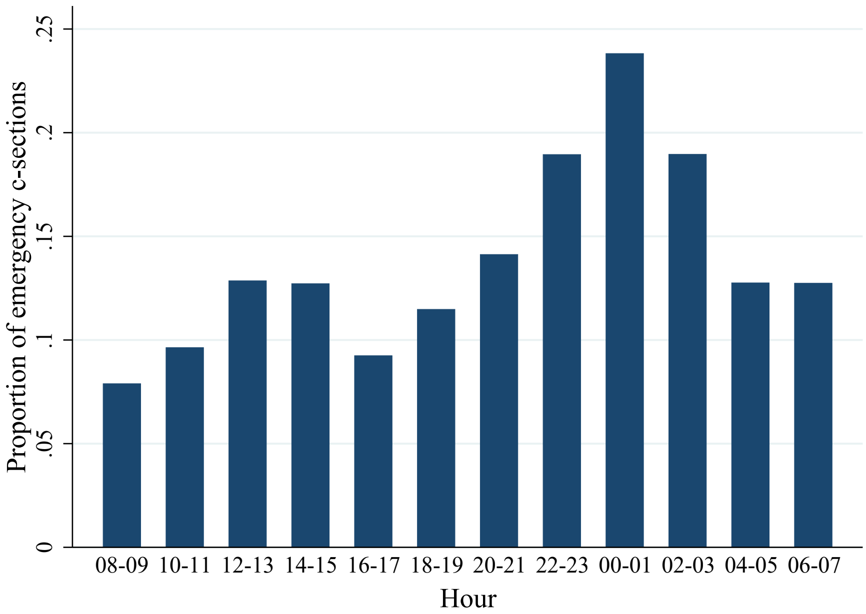


Figure 1: Proportion of unplanned c-sections by time of day

3.3 Identification strategy

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

Table 1: Maternal Characteristics by Type of Birth

	Means		p-value
	Eutocic birth	C-section	for difference
A. Personal characteristics			
Mother's age	31.466	32.828	0.000
Level of education			
No school	0.037	0.022	0.044
Primary school	0.278	0.206	0.000
Secondary school	0.502	0.609	0.000
University education	0.182	0.164	0.234
Non-Spanish	0.278	0.199	0.000
Single	0.017	0.015	0.602
Mother's weight	65.471	67.830	0.000
Mother's height	1.653	1.595	0.547
B. Pregnancy characteristics			
Tobacco during pregnancy	0.119	0.134	0.256
Alcohol during pregnancy	0.003	0.007	0.067
Gestation weeks	39.267	38.863	0.000
Previous c-section	0.064	0.223	0.000
Obstetric risk	0.350	0.580	0.000
Intrapartum pH	7.296	7.245	0.000
Birthweight	3290.334	3181.038	0.000
Induction	0.189	0.431	0.000
Observations	4201	685	4886

We define a binary variable CS_i equal to one if the mode of delivery is an emergency c-section and zero if it is an eutocic delivery, i.e. a vaginal delivery with no interventions. Infant health H_i refers to either Apgar scores or umbilical cord pH. We would thus like to estimate the following equation:

$$H_i = \beta_0 + \beta_1 CS_i + \beta_2 X_i + \epsilon_i \tag{1}$$

where X_i is a set of covariates that include information on mothers' personal and pregnancy characteristics. As discussed earlier, the estimation of equation (1) is, however, likely to provide biased estimates of β_1. To overcome this potential endogeneity, we use an IV approach, instrumenting the type of birth with an indicator for the time of day the infant is born. Therefore, our first stage is as follows:

$$CS_i = \gamma_0 + \gamma_1 earlynight_i + \gamma_2 X_i + v_i \quad (2)$$

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

The identifying assumption is that $earlynight_i$ is not correlated with ϵ_i, but this assumption entails two conditions. The first is that the instrument is as good as randomly assigned. We provide suggestive evidence that this is the case by comparing personal and pregnancy characteristics of mothers who give birth between 11 pm and 4 am and those during the rest of the day in Table 2. Mothers are similar with respect to their educational level, weight and height, alcohol and tobacco consumption habits during pregnancy, gestational length, obstetric risk, weight of the newborn, or previous c-sections. The level of intrapartum pH, a measure of fetal distress during labor – a major cause of emergency c-sections – is also equivalent (see more on this in Section 5.3). We find some slight differences between mothers across time of day with respect to nationality (there are slightly more non-Spanish women during the day shift) and marital status (more unmarried women during the day). However, these differences are very small in magnitude. We also find that the proportion of women whose labor was induced is higher during the early hours of the night (26.1%) compared to the rest of the day (21.2%). This is something one might expect from our institutional setting, since in the hospitals in our sample most inductions are performed in the morning and, given the average duration of labor, these women are more likely to give birth during the early hours of the night. We control in our main specification for all of these differences and perform a robustness check excluding inductions in Section 5.2, where we find that our conclusions still hold. Overall, we thus feel confident with the assumption that there is no selection of women into the different shifts that could threaten our identification.

Additionally, identification requires the exclusion restriction to hold; that is, the instrument should affect infant health only through the increased probability of having a c-section. One potential concern is that the quality of medical care could change depending on the time/shift. In order to overcome this problem, as a robustness check we perform the analysis using variation in the probability of having a c-section only during the night, thus holding the quality of

Table 2: Maternal Characteristics by Time at Delivery

	Means		p-value
	Rest of the day	Early night	for difference
A. Personal characteristics			
Mother's age	31.592	31.883	0.120
Level of education			
No school	0.037	0.028	0.181
Primary school	0.267	0.271	0.817
Secondary school	0.517	0.518	0.943
University education	0.179	0.183	0.779
Non-Spanish	0.275	0.237	0.012
Single	0.019	0.009	0.024
Mother's weight	65.698	66.164	0.355
Mother's height	1.655	1.608	0.556
B. Pregnancy characteristics			
Tobacco during pregnancy	0.120	0.125	0.679
Alcohol during pregnancy	0.003	0.005	0.481
Gestation weeks	39.209	39.215	0.923
Previous c-section	0.084	0.095	0.228
Obstetric Risk	0.379	0.394	0.394
Intrapartum pH	7.274	7.286	0.337
Birthweight	3276.399	3270.175	0.728
Induction	0.212	0.261	0.001
Observations	3796	1090	4886

medical care constant (see section 5.1).

4 Results

Tables 3 and 4 present the results for the OLS estimation of equation (1) for the different measures of neonatal health. In table 3, the first column for each outcome presents the results without controls, the second column incorporates controls for maternal characteristics, and finally the third column adds information about the pregnancy. All specifications include hospital and weekday fixed effects, the sample is restricted to single births, and we cluster standard errors at the hospital-day level[3]. The results show that delivering via a c-section is associated with a significant decline of Apgar scores 1 and 5 and with a lower probability of having moderate

[3] All estimations hereafter use clustered standard errors at the hospital-day level. We show in Tables A2 and A3 in the appendix that our IV results are robust to alternative standard error estimations.

pH, but not of having severe pH. Table 4 presents the results for other outcomes of neonatal health. As it can be seen, babies born by cesarean section are more likely to need reanimation and to go to the Intensive Care Unit, but they are also less likely to die.

Table 3: OLS Results – Neonatal Health

Panel A. Apgar Scores	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Emergency CS	-0.590***	-0.586***	-0.488***	-0.236***	-0.238***	-0.156***
	(0.058)	(0.058)	(0.064)	(0.039)	(0.038)	(0.047)
Mean of Y		8.945			9.809	
Observations		4886			4884	
Panel B. pH Level	pH $<$ 7.2			pH $<$ 7.15		
	(1)	(2)	(3)	(1)	(2)	(3)
Emergency CS	-0.057***	-0.058***	-0.070***	0.002	0.001	-0.010
	(0.019)	(0.020)	(0.021)	(0.015)	(0.015)	(0.016)
Mean of Y		0.215			0.098	
Observations		3758			3758	
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

As explained above, these estimates are likely to be biased because mothers giving birth by c-section and vaginally are not comparable, and because we cannot identify which kind of c-section is driving the results. The results for the IV estimation of the effects of non-medically indicated c-sections on Apgar scores 1 and 5 are shown in Table 5. The first stage F-statistics are larger than 39 for the different specifications, so following Stock and Yogo (2005) critical values with one endogenous variable and one IV (16.38), we can reject the null hypothesis that our instrument is weak. In line with our descriptive analysis, Panel B shows that births that take place between 11 pm and 4 am are around 8 percentage points more likely to be by cesarean.

In the first row of the table below (Panel A), we observe that a c-section has a negative impact on both Apgar Score 1 and Apgar Score 5. The estimated effects are large and significant. In the specification with the full set of controls (column 3), an emergency c-section reduces Apgar Score 1 by 1.161 points. This effect is larger than one standard deviation (1.117) and is significant at the 5% significance level. An emergency c-section also has a negative impact on Apgar Score 5. In this case the coefficient is -0.942, and again, is larger than one standard

Table 4: OLS Results – Other Outcomes

	Intensive Care Unit		Reanimation		Exitus	
	(1)	(2)	(3)	(4)	(5)	(6)
Emergency CS	0.143***	0.112***	0.095***	0.077***	-0.002	-0.007*
	(0.016)	(0.014)	(0.014)	(0.014)	(0.002)	(0.004)
Mean of Y	0.057		0.073		0.005	
Observations	4886	4886	4886	4886	4886	4886
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls		✓		✓		✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

deviation (0.818) and significant at the 5% significance level.

Most of the newborns in our sample have an Apgar score 1 equal to 9 and an Apgar score 5 equal to 10 (see figure A6). We thus perform a similar analysis but using as dependent variable an indicator for having Apgar scores 1 and 5, respectively, lower than 10 (table A4), and both scores lower than 9 (table A5). Our qualitative conclusions hold, as we find that a non-medically justified c-section, as compared to an eutocic delivery, increases the probability of having Apgar scores 1 and 5, respectively, below 10 by around 30% and 40%, and the probability of having Apgar scores 1 and 5 below 9 by 40% and 17%. Finally, Figure A7 in the appendix provides an overview of the size of the coefficients for different thresholds of Apgar 1 and 5, respectively, as dependent variables. This is relevant, since decreases in Apgar scores are non-linearly related to the health of the newborn. We see that non-medically justified c-sections significantly increase the probability of having Apgar scores lower than 10, 9, and 8, but not lower than 7 or inferior levels. Therefore, these marginal c-sections increase the probability of deviating from the perfect scores, which are the mode in our sample, but we do not see significant effects in the left tail of the distribution.

In Table 6 we estimate the impact of a c-section on the probability of the pH level being below different thresholds: pH levels below 7.2 (low pH) and pH below 7.15 (very low pH). As can be seen, a c-section increases the probability of both indicators and the coefficients are significant for all the specifications, at the 10% significance level for low pH and at the 5% significance level for very low pH. In particular, a c-section increases the probability of low and very low pH by approximately 45 percentage points.

We also perform the same analysis for other infant health outcomes. Results can be found in

Table 5: IV Estimation – Apgar Scores

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

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7. Although we might expect an effect on needing intensive care, reanimation, or neonatal mortality, we only observe a slightly significant impact on intensive care unit utilization in one specification. However, the estimates are large in magnitude, so we cannot rule out an effect on these outcomes.

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

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

Table 6: IV Estimation – pH Level Indicators

	pH < 7.2			pH < 7.15		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Emergency CS	0.408*	0.417**	0.451*	0.406**	0.413**	0.445**
	(0.211)	(0.211)	(0.234)	(0.163)	(0.164)	(0.180)
Mean of Y		0.215			0.098	
Panel B. First stage						
Early night	0.085***	0.085***	0.077***	0.085***	0.085***	0.077***
	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)
Observations	3751	3751	3751	3751	3751	3751
First-stage F	30.979	31.092	29.505	30.979	31.092	29.505
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

However, if we compare the results for the other outcomes (see tables 4 and 7), we observe that in this case the coefficients for the OLS are larger and significant: c-sections are associated with an increased probability of needing intensive care and reanimation, but with a reduction in neonatal mortality. This suggests that these medically-indicated c-sections are performed in order to assist infants in distress who need immediate support. On the other hand, the IV estimates are not significant, arguably because the effects of non-medically indicated c-sections are short-lived: in spite of the worsening in Apgar scores and pH, we do not find substantial evidence that these negative effects translate into needing intensive care, reanimation, or increased mortality risk.

5 Robustness checks

5.1 Exclusion restriction: variation within the night shift

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

Table 7: IV Estimation – Other Outcomes

	Intensive Care Unit		Reanimation		Exitus	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. 2SLS						
Emergency CS	0.161*	0.137	0.109	0.089	0.030	0.028
	(0.094)	(0.100)	(0.100)	(0.114)	(0.030)	(0.034)
Mean of Y	0.057		0.073		0.005	
Panel B. First stage						
Early night	0.088***	0.078***	0.088***	0.078***	0.088***	0.078***
	(0.013)	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)
Observations	4886	4886	4885	4885	4886	4886
First-stage F	43.974	39.192	43.959	39.079	43.974	39.192
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls		✓		✓		✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to a reduction in the quality of care during this time.

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

Results for the IV estimation using variation during the night can be found in Tables 8 and 9. Despite the smaller sample size, we again find that an emergency c-section reduces both Apgar scores 1 and 5 and increases the probability of having a pH lower than 7.2 and 7.15. The coefficients remain large and significant at the 5% significance level. We interpret these results as evidence in favor of our exclusion restriction.

Table 8: IV Estimation – Apgar Scores during the Night

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Emergency CS	-1.445**	-1.476**	-1.439*	-1.235**	-1.261**	-1.293*
	(0.708)	(0.743)	(0.861)	(0.566)	(0.593)	(0.679)
Mean of Y		8.919			9.793	
Panel B. First stage						
Early Night	0.067***	0.065***	0.055***	0.067***	0.065***	0.055***
	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)
Observations	2553	2553	2553	2552	2552	2552
First-stage F	19.759	18.243	14.792	19.665	18.138	14.724
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births and to births taking place between 8pm and 8am. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.2 Excluding inductions

The comparison of maternal characteristics in Table 2 showed that mothers giving birth in the early hours of the night are more likely to have had their labor induced. Inductions can be scheduled, typically because the pregnancy has gone beyond full term and labor has not spontaneously started, or can be unscheduled if the mother's waters break but labor does not begin (NICE, 2008). If an induction is to be scheduled, the hospitals in our sample usually plan the latter for the morning, such that after progression of labor at average pace these women are expected to give birth in the evening or during the early hours of the night.

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

Table 9: IV Estimation – pH Level Indicators during the Night

	pH < 7.2			pH < 7.15		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Emergency CS	0.979**	1.038**	1.162**	0.779**	0.829**	0.922**
	(0.431)	(0.459)	(0.521)	(0.313)	(0.333)	(0.378)
Mean of Y		0.209			0.097	
Panel B. First stage						
Early Night	0.061***	0.059***	0.053***	0.061***	0.059***	0.053***
	(0.017)	(0.017)	(0.016)	(0.017)	(0.017)	(0.016)
Observations	1961	1961	1961	1961	1961	1961
First-stage F	12.238	11.325	10.466	12.238	11.325	10.466
Maternal controls		✓	✓			✓
Pregnancy controls			✓		✓	✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births and to births taking place from 8pm to 8am. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

be more sensitive to external factors and incentives.

In any case, even if the decision to perform a c-section on mothers with induced labor was more dependent on doctors' routines or incentives than on the health conditions of the mother and the baby, if our analysis was driven by this type of mother alone, we would not be able to disentangle the effect of c-sections from the effect of medical inductions. In our main specifications we directly control for whether labor was induced, but in Table 10 we also repeat our analysis excluding inductions from our sample. Here we see that, despite the reduction in the number of observations, our qualitative conclusions hold: births in the early night are still more likely to end up as cesarean sections, and these have a negative and significant impact on Apgar scores. We thus conclude that, although inductions seem to make our first stage stronger as they might offer room for discretionary behavior, our findings do not depend on including them.

5.3 Emergency c-sections: medically indicated versus non-medically indicated

In order to ensure that the health effects we find are not due to medically indicated c-sections, we explore whether the c-sections captured by our instrument are correlated with the same

Table 10: Robustness Check – Excluding Inductions

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Emergency CS	-2.271**	-2.312**	-2.430**	-1.905**	-1.972**	-2.073**
	(1.102)	(1.147)	(1.183)	(0.935)	(0.982)	(1.013)
Mean of Y		9.001			9.841	
Panel B. First stage						
Early Night	0.043***	0.042***	0.041***	0.043***	0.042***	0.041***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Observations	3795	3795	3795	3793	3793	3793
First-stage F	10.801	10.282	10.762	10.748	10.222	10.668
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births and excludes inductions. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk and gestation weeks. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

indications that should predict a medically-necessary cesarean section.

One of the main medical indications for an emergency c-section is fetal distress. This is monitored during labor by several means, including watching the cardiac frequency or measuring the fetal scalp pH. Similar to the umbilical cord pH, if the fetal scalp pH is too low (namely, below 7.20) it suggests that the fetus is not getting enough oxygen. If this situation persists for too long, it could threaten the baby's health and the clinical advice is to perform an emergency c-section. Therefore, while medically-indicated c-sections should be predicted by fetal distress, those that are not medically-indicated, but performed out of convenience, should not.

We would not, a priori, expect our instrument to be correlated with fetal distress: there is no apparent reason why births at night should present higher risks for the fetus. The distribution of the intrapartum pH across a 24-hour cycle seems to confirm this (see Figure A8): we see a uniform distribution throughout all hours of the day, suggesting that there are no systematic differences in average fetal distress across time. We can also test for this formally, although we only have information about fetal scalp pH for a small part of our sample (around 200 observations). We do this in Table 11. Columns (1) and (3) present the results of regressing the dummy for all emergency c-sections on the level of intrapartum pH and on an indicator for low intrapartum pH (below 7.2), respectively. We can see that lower levels of pH are strongly associated with a higher probability of performing a c-section, and that having an intrapartum pH below 7.2 is also associated with a higher probability of having a c-section. On the other hand, in columns (2) and (4) we perform the same analysis but substitute the dependent variable for the predicted c-sections from our first stage – that is, a variable keeping only the variation in the probability of having a c-section that is predicted by our instrument. In this case, we do not see any significant correlation with the two measures of intrapartum pH. Therefore, the c-sections captured by our instrument do not seem to be predicted by fetal distress but by other reasons. We interpret this as supporting evidence that the negative health impacts that we find are due to non-medically indicated cesarean sections.

Table 11: Robustness Check: Fetal Distress and C-Sections

	(1) Emergency CS	(2) Predicted CS	(3) Emergency CS	(4) Predicted CS
Intrapartum pH	-1.702***	0.037		
	(0.360)	(0.030)		
Intra. pH < 7.2			0.309***	-0.008
			(0.085)	(0.006)
Observations	216	216	216	216

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. Sample is restricted to single births. All specifications include hospital and weekday fixed effects, and the full set of maternal and pregnancy controls. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.4 Doctors' leisure incentive: some suggestive evidence

Although it is not crucial for our identification strategy, in this section we endeavor to shed some light on the mechanism behind the exogenous variation in the probability of a c-section at different times of the day that we observe in our data.

As mentioned previously, the most plausible explanation is that doctors have a higher incentive to perform c-sections in the early hours of the night as, at this time, the opportunity cost of time becomes more salient. This is because doctors have already been working for more than 12 hours and if they perform the c-section and do not have other mothers to attend, they can rest for the remainder of their shifts. Accordingly, we would expect that doctors are more likely to perform a non-medically indicated c-section on nights when there is only one birth compared to nights when there is more than one ongoing delivery.

We provide suggestive evidence that this is the case. The first column in Table 12 shows the first stage coefficient for nights when only one delivery took place and the second column for nights with more than one birth. In line with our argument, the results of this exercise suggest that doctors perform more non-medically indicated c-sections in the early hours of the night when they have only one ongoing delivery.

Table 12: First Stage: Busy vs. Non-Busy nights

	(1) Single-birth nights	(2) Multiple-birth nights
Early Night	0.106***	0.069***
	(0.026)	(0.015)
Observations	1252	3152

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. Sample is restricted to single births. Single-birth nights are defined as days in which there is only one delivery from 8 pm to 8 am, whereas multiple-birth nights are those in which more than one delivery occurs during these hours. All specifications include hospital and weekday fixed effects, maternal and pregnancy controls. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 Conclusions

This paper provides new evidence of the adverse effects of avoidable cesarean sections on newborn health. In order to overcome potential omitted variable bias and abstract from those cases in which c-sections respond to a clear clinical indication, we make use of a novel instrument that exploits variation in the probability of receiving a c-section that is unrelated to maternal and fetal health: variation in time of birth.

Our results suggest that these non-medically indicated c-sections lead to a significant worsening of two frequent measures of newborn health: Apgar scores and the pH of the umbilical cord. In particular, the deterioration in these outcomes likely captures increased respiratory problems related to the presence of amniotic liquid in the newborn's lungs. The relative decline in Apgar scores might also capture reduced excitability and muscle tone. All in all, these findings are consistent with the medical literature, which has identified vaginal delivery as a crucial programming event in the baby's life (Hyde et al., 2012).

Although the size of the effects we find is of statistical and medical significance – declines range between 1 and 1.5 standard deviations for all neonatal health outcomes – we do not find evidence that these effects translate into a significant increase in the need for reanimation or intensive care or into increased risk of neonatal death. Therefore, the effects we find might not be overly severe or may fade over time. Nonetheless, neither do we find evidence of any health benefits of these non-medically justified interventions. More research is needed in order to obtain a more complete understanding of the causal effect of non-medically indicated c-sections on the health of the infant and the mother in the longer-run. Given, however, the monetary cost of these avoidable interventions (the average cost of a c-section for the Spanish public health system is around 1.8 times that of a vaginal delivery[4]), the absence of health benefits, and significant health costs, policies aimed at avoiding excessive use of this procedure are likely to increase efficiency.

Similarly, more work is needed to understand the decisions of doctors driving the observed time variation in c-section rates. In this paper we have only provided some suggestive evidence of the mechanism behind this variation, which is consistent with the findings of previous studies. Our results point to the need to revise the incentives created by the shift structure and long working hours of physicians, so as to reduce avoidable interventions.

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

References

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

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

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

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

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

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

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

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

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

Gao, C., Yuan, L., and Wang, J. (2009). Role of pH value of umbilical artery blood in neonatal asphyxia. *Chinese Journal of Contemporary Pediatrics*, 11(07):521.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix

Figure A1: Distribution of Umbilical Cord pH by Levels of Apgar 1 and 5

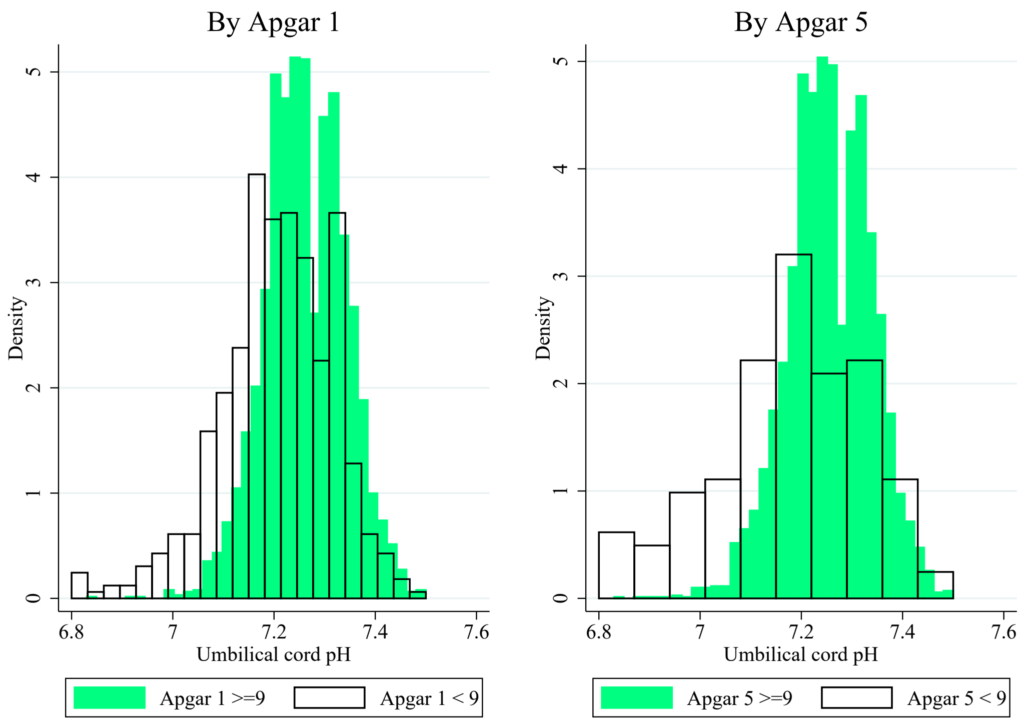
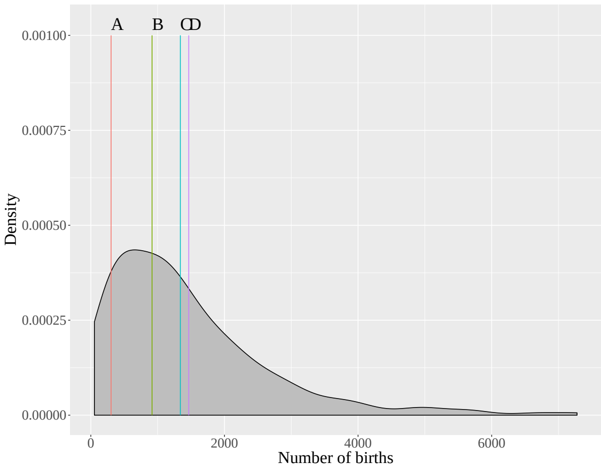
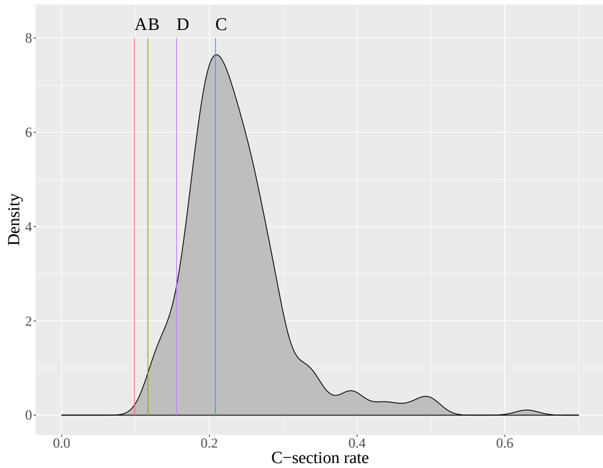


Figure A2: Distribution of Number of Births and C-Section Rates in all Spanish Public Hospitals (Estadística de Centros Sanitarios de Atención Especializada 2013) Compared to Hospitals in our Sample in 2015

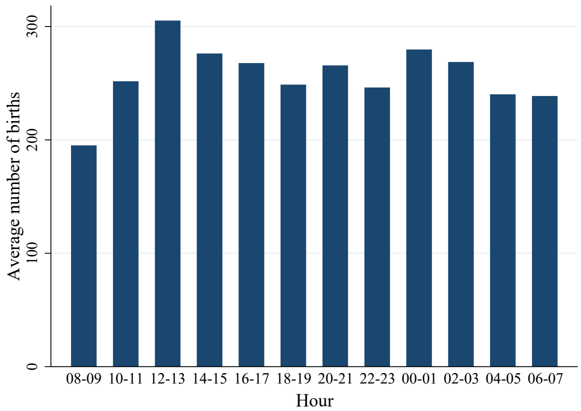


(a) Number of Births Attended in a Year

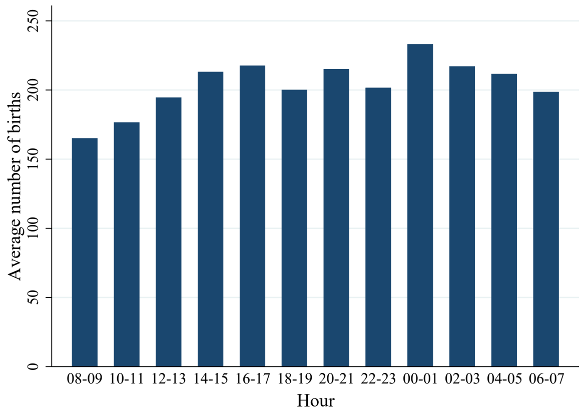


(b) C-Section Rate in a Year

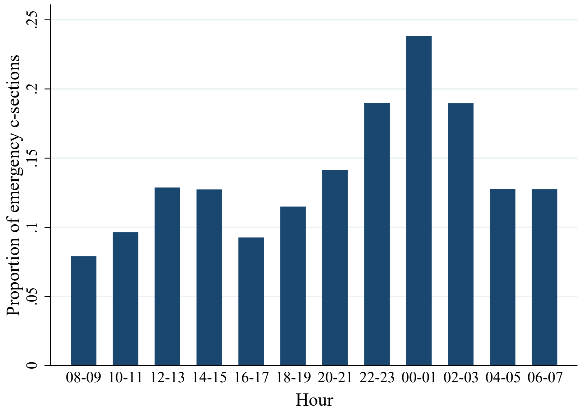
Figure A3: Distribution of Different Types of Births across Times of Day



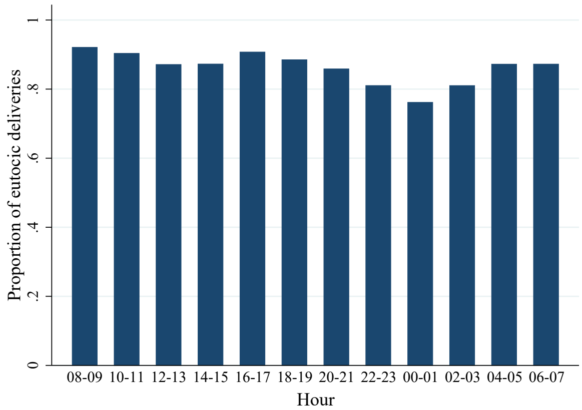
(a) All Births



(b) Emergency C-Sections and Eutocic Births



(c) Emergency Cesarean Sections



(d) Eutocic Deliveries

Notes: These figures represent the distribution of different types of births across times of day, grouped by intervals of two hours. Figure (a) represents the number of births per two hours using the full sample of 6,163 observations. Figures (b)-(d) use our usual sample of 4,886 observations. Figure (b) shows the number of births per two hours in this restricted sample, which includes only emergency c-sections or eutocic births. Figure (c) represents the proportion of emergency c-sections over the total number of births of this sample, while figure (d) displays the proportion of eutocic deliveries.

Figure A4: Proportion of Emergency C-Sections by Physicians' Hours Worked (Loess Estimate)

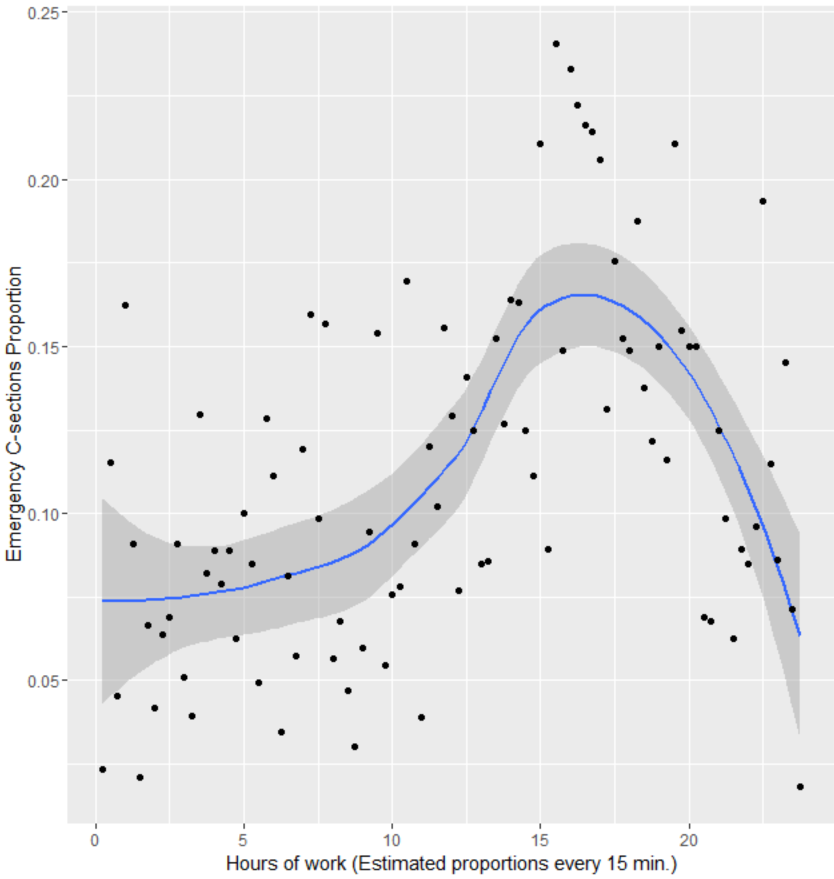


Figure A5: Predicted Probability by Doctor of Attending Births during the Early Night

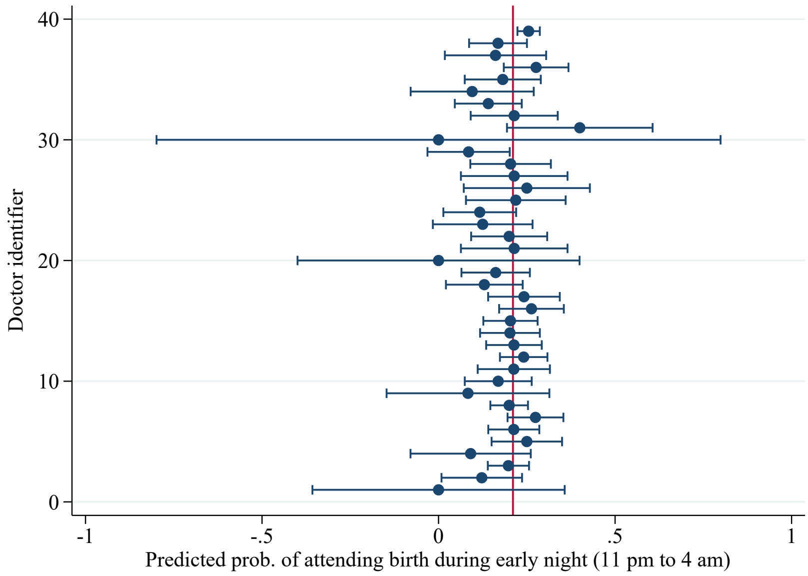
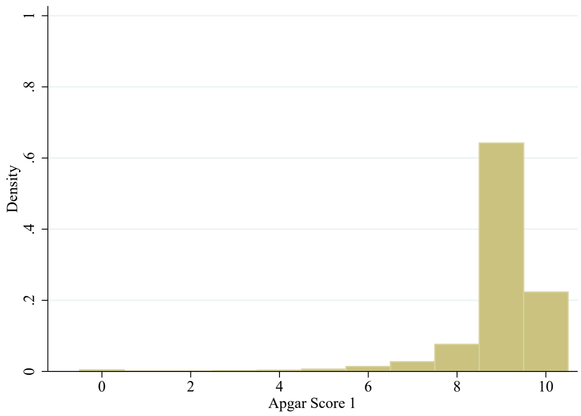
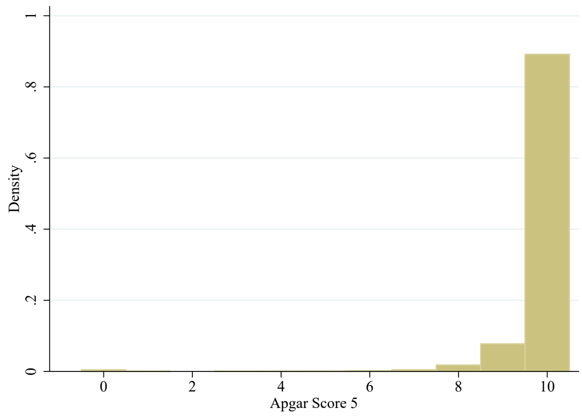


Figure A6: Distribution of Apgar Scores

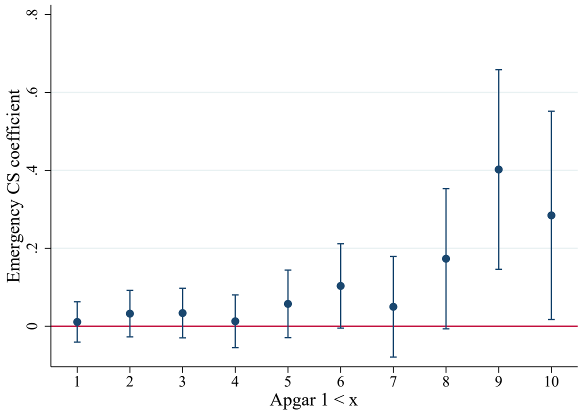


(a) Apgar Score 1

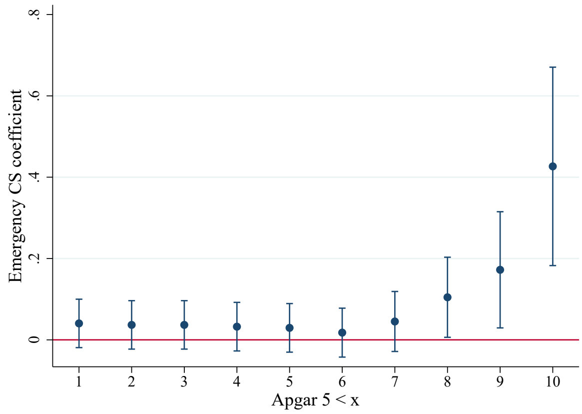


(b) Apgar Score 5

Figure A7: The Effect of Non-Medically Justified C-Sections: IV Coefficients by Apgar Threshold



(a) Apgar Score 1



(b) Apgar Score 5

Figure A8: Average Intrapartum pH by Time of Day

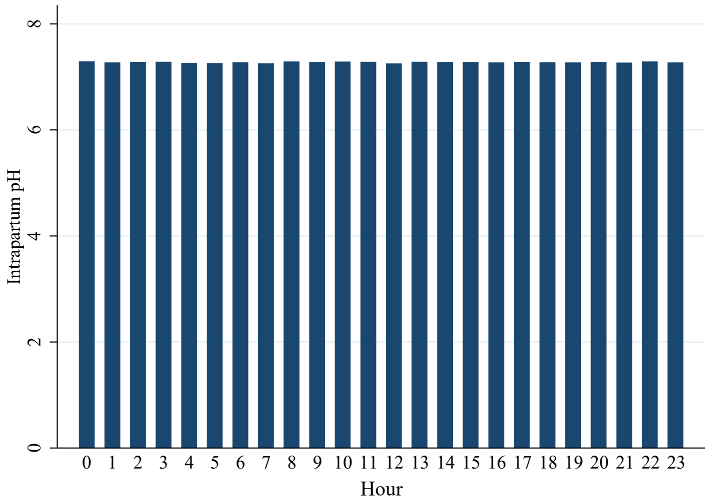


Table A1: Summary Statistics

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

Table A2: IV Estimation – Apgar Scores: Standard Errors Robustness

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Emergency CS	-1.161**	-1.161**	-1.161**	-0.942**	-0.942**	-0.942**
	(0.521)	(0.514)	(0.516)	(0.423)	(0.426)	(0.433)
Observations	4886	4886	4886	4884	4884	4884
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls	✓	✓	✓	✓	✓	✓
Cluster (day)	✓			✓		
Cluster (day-hospital)		✓			✓	
Robust			✓			✓
Mean of Y	8.945	8.945	8.945	9.809	9.809	9.809

Notes: All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3: IV Estimation – pH Level Indicators: Standard Errors Robustness

	pH < 7.2			pH < 7.15		
	(1)	(2)	(3)	(1)	(2)	(3)
Emergency CS	0.451*	0.451*	0.451*	0.445**	0.445**	0.445**
	(0.237)	(0.234)	(0.238)	(0.177)	(0.180)	(0.185)
Observations	3751	3751	3751	3751	3751	3751
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls	✓	✓	✓	✓	✓	✓
Cluster (day)	✓			✓		
Cluster (day-hospital)		✓			✓	
Robust			✓			✓
Mean of Y	0.215	0.215	0.215	0.098	0.098	0.098

Notes: All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4: IV Estimation – Apgar Score < 10

	Apgar Score 1 <10			Apgar Score 5 <10		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Emergency CS	0.296**	0.298**	0.285*	0.404***	0.420***	0.427***
	(0.143)	(0.145)	(0.164)	(0.125)	(0.128)	(0.144)
Mean of Y		0.777			0.109	
Panel B. First stage						
Early night	0.090***	0.088***	0.078***	0.090***	0.088***	0.078***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	4886	4886	4886	4884	4884	4884
First-stage F	45.329	43.974	39.192	45.222	43.852	39.102
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: IV Estimation – Apgar Score < 9

	Apgar Score 1 <9			Apgar Score 5 <9		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Emergency CS	0.383***	0.403***	0.402***	0.160**	0.167**	0.172**
	(0.135)	(0.138)	(0.156)	(0.075)	(0.077)	(0.087)
Mean of Y		0.135			0.031	
Panel B. First stage						
Early night	0.090***	0.088***	0.078***	0.090***	0.088***	0.078***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	4886	4886	4886	4884	4884	4884
First-stage F	45.329	43.974	39.192	45.222	43.852	39.102
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: Standard errors (in parentheses) are clustered at the hospital-day level. All specifications include hospital and weekday fixed effects. Sample is restricted to single births. Maternal controls include: level of education, nationality, maternal weight, height, age and marital status. Pregnancy controls include: previous c-section, prenatal care, obstetric risk, gestation weeks and induced labor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Universitat Pompeu Fabra
Barcelona

CRES
CENTRE DE RECERCA EN ECONOMIA I SALUT