Essays in Health Economics

Gabriel Alejandro Facchini Palma

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 02 October 2017

European University Institute
Department of Economics

Essays in Health Economics

Gabriel Alejandro Facchini Palma

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Prof. Andrea Ichino, EUI, Supervisor
Prof. Jérôme Adda, EUI and Bocconi University
Prof. Joseph Doyle, MIT Sloan School of Management
Prof. Libertad González, Universitat Pompeu Fabra

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I Gabriel Alejandro Facchini Palma certify that I am the author of the work Essays in Health Economics I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

The copyright of this work rests with its author. [quotation from it is permitted, provided that full acknowledgement is made.] This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

Signature and Date: 21/09/2017

Abstract

This aims at better understanding the drivers behind the volume-outcome relationship found in many studies in the medical and health-economics literature.

In the first chapter I investigate the relationship between workload and choice of treatment. Using detailed microdata on childbirth, I exploit a quasi-random assignment of patients attempting to have a natural delivery to different ratios of patients-to-midwives and compare their likelihood of changing delivery method. I find that women who face a ratio higher than 1.33 are 34% more likely to give birth by cesarean section (C-sections). This effect is larger for patients who were already admitted with a higher risk of C-section, since provision of proper and timely care matters more for these patients. Because C-sections are faster than vaginal deliveries, the medical team may find it appealing to do more C-sections when time constrained. Using civil status as a proxy for bargaining power -assuming single women are on average more likely to be alone-, I find that only single patients are subjected to unnecessary surgery.

The second chapter documents the existence of 'learning-by-doing' effects in physicians' performance. More specifically, I test whether cesarean-section surgeons who have performed more procedures in the recent-past observe an improvement in performance. By using data from the Italian health care system, where patients are not allowed to choose a physician, I eliminate concerns regarding possible bias from selective referral -a problem in previous studies. Using four years of birth certificates data from one large hospital I find that, for emergent cases, performing one additional procedure reduces the likelihood of neonatal intensive care unit admission by nearly 1.2 percentage points (5.5%) and of being born with a low Apgar Score by about 1.1 percentage points (10%), all else equal.

Acknowledgments

All good things come to an end. I feel extremely lucky to have been given the opportunity to spend the last five years of my life doing my dissertation at the EUI. The completion of this thesis was a challenging and exciting journey that I was possible thanks to the many people I am highly indebted to. I will briefly mention a few of them, but I am extremely fortunate as to have many more in my mind. If you are reading this you are most likely one of them, so THANK YOU!

I would like to extend my gratitude to Andrea Ichino and Jérôme Adda for their invaluable support and advice throughout my doctoral studies. During this time I benefited greatly from discussions with Juan Dolado, Daniela Iorio, Matilde Machado and Matthew Neidell. My gratitude goes to Joseph Doyle and Liertad González as well, for accepting to be part of my committee and for their valuable comments which will improve the essays this thesis is comprised of. This thesis would have not been possible without all the comments and discussions during our meetings in the "Micro-econometrics working group" and the "Apaleados working group".

I am also very thankful to all the faculty members of the Economics Department at the EUI, and to the staff members of VLF/VSP. A special thank you goes to Arpad, Sarah, Jessica, Lucia, Lori, Sonia and Anto for their support and empathy with us students.

I am extremely grateful to the personnel at the maternity ward of Azienda Ospedaliero Universitaria Careggi for their collaboration and patience while I was trying to understand the functioning of the ward. The same goes to all my friends who are medical doctors (a.k.a. real doctors), and would take the time to talk with me even though they could not understand why an economist could be interested on these issues.

During the time in Florence I also discovered I can play football! Playing at least once a week kept me grounded ("mi cable a tierra"). For this I thank my teammates in the Hurrikeynes and the Squadra Fantastica.

The last five years would not have been the same without the valuable support, discussions and company of my friends and colleagues, Chaco, Luis, Lian, Pablo, Gaby, Ine (l@s vag@s), Benni, Rasmus, Filip (TSFTAoRC), Toni, Andresa, Ricca and many more.

A special thanks goes to Chaquito, my friend since the first moment I arrived in Florence (and before?), even though we are two very different people. You were always there ready to lend a sympathetic ear. I hope we enjoy more adventures in the future, be it for tapas in Spain, or a parripollo in Rio.

Finally, this thesis is dedicated to my family, whose unconditional support and endless encouragement have made this possible. To my parents, Marcos and Nelida, I hope to become half as good of a professor as you both are. To my brothers, Mauri and Rodri, with whom I could share so many adventures. I always felt reassured knowing they were there for me despite the distance. Above all I would like to thank Mariana for her support and patience, because she has endured all my ups and downs during these past years and stayed nevertheless.

To staying curious...

Contents

Chapter 1

Low Staffing in the Maternity Ward: Keep Calm and Call the Surgeon

1.1 Introduction

Over the last decades health care systems in developed countries have been under constant pressure to reduce costs, despite facing an increasing demand for health care services. In order to avoid a trade-off between cutting down on costs and a negative impact on patients' health outcomes, experts currently point towards the reduction of waste as the best way to go.[1] Among the several sources of waste, two widely cited ones are the lack of adoption of known best practices (e.g. effective preventive care) and overtreatment, that is, the carrying out of treatments that cannot possibly improve the patients' health (e.g. cases of physician induced demand). These two sources of waste are particularly salient in maternity ward settings.

The role of midwives -as opposed to physicians- in assisting birth speaks to the first point. Whereas relevant public health authorities have recently recognized that midwife-led care during labor is safer for low-risk pregnancies[2], the media and midwifery colleges have long spoken of a "shortage of midwives"[3], which was also acknowledged by the World Health Organization (WHO) in 2009[4]. At the same time, cesarean sections (C-sections) rank high among greatly overused

[1]See, for example, Berwick and Hackbarth (2012).

[2]For example, the National Institute for Health and Care Excellence (NICE) updated its guidelines in this direction in 2014.

[3]In a 2015 report, The Royal College of Midwives estimates that the UK "...needs 2,600 more midwives to be able to cope with the number of births the country is experiencing...". The Federal Association of Midwives of Spain (FAME) has as main objective to address the shortage of midwives in the health care system. The president of the Italian Midwifery Association recently stated that "...there is a shortage of midwives. Too few to guaranty the proper level of care that other European Countries have".

[4]**?**.

interventions[5], and governments and clinicians have expressed concern about its potential negative impact on patients' health[6]. Indeed, C-sections not only cost more than vaginal deliveries, but they also imply higher risks for both mother and infant[7] and, according to a growing medical literature, are associated to lower long-term outcomes of children's health[8]. In addition, because vaginal delivery after a C-section (VBAC) is very unlikely[9], one C-section sets a path dependency for more C-sections in future births. There is also evidence that women who follow a C-section are more likely to have less children[10], something that is particularly alarming in developed countries with already low fertility rates.

In light of these concerns, a natural question is whether a situation of low staffing can result in more unnecessary C-sections being performed. This can happen either as a direct consequence of high workload -with midwives devoting less time to each patient, which increases the probability of complications that lead to surgery- or because physicians may find it optimal to induce some patients towards a C-section independently of their health status. Since a C-section takes less time than a vaginal birth -no need to wait for the appropriate dilation of the cervix-, midwives' workload can be reduced by redirecting patients to the operative theater.

This study causally tests whether patients follow a different delivery method depending on the effective staff level in the maternity ward at the moment of admission. It exploits a simple natural experiment: the majority of patients follow the natural course of birth and only go to the hospital once labor has already started and/or their water has broken (unlike, for example, scheduled cesarean sections). The effective staff level (e.g. the staff per patient ratio) observed by these patients at admission is orthogonal to their demographic and health characteristics (and to their ex-ante probability of delivering by C-section). The effective staff level at admission changes with the number of patients who arrived before and the number of midwives present in the delivery room, two variables that are unknown for the incoming patient.

The data for this project comes from a census of births from a large public hospital in Italy for the period 2011-2014. Three features of this dataset make it well suited for tackling the issue at hand. First, birth certificates have precise information on delivery method, allowing the identification of scheduled and unscheduled patients. Second, using patient's ID, each certificate was merged

[5]While the international healthcare community considers an ideal rate of C-sections to be between 10-15%, country average rates in Europe vary from as low as 15.6% in The Netherlands to as high as 36.8% in Italy (OECD data 2012).

[6]WHO Statement on Caesarean Section Rates, WHO (2015).

[7]See Deneux-Tharaux et al. (2006); Gregory et al. (2012); Curtin et al. (2015).

[8]Infants born by C-section are not exposed to the maternal bacteria of the birth canal and as a consequence have different intestinal bacteria, which can affect their immune system and other important processes. For a meta-analysis of this literature see (Blustein and Liu, 2015).

[9]VBAC rate is only 8.3% in the US, and 12% in Italy.

[10]Norberg and Pantano (2016).

with hospital administrative data containing the exact time of admission and discharge. I use this information to compute the actual number of patients in the delivery room at each point in time. Finally, this is complemented with data on the number of midwives scheduled by month, day of the week and shift.

Results suggests that there is a non-linear relationship between effective midwifery staff and delivery method: a newly admitted patient who faces a ratio of patients-to-midwives higher than 1.33 is 34% more likely to give birth by C-section. This means that, for first-time mothers, about 1.2 p.p. (or 5.7%) of all C-sections (both scheduled and unscheduled) are the consequence of low midwifery staffing.

The second part of the analysis looks at possible mechanisms behind this change in delivery method. One possibility is that, in situations with a high ratio of patients-to-midwives, the time dedicated to each patient is lower and the quality of care inappropriate, eventually resulting in the need for C-section. If that is true, then one should see patients with marginally lower health being more affected. In order to test this hypothesis, two types of patients are compared: a low-health type, formed by those patients who had an emergency visit during their pregnancy or whose babies had an extreme weight at birth, and a high-health type, with all the remaining patients. Indeed, the gap between the probability of having a C-section between a low-health and a high-health patient widens with a higher workload.

Another factor that can explain the rise in C-sections alongside with workload is the presence of physician induced demand (PID). Because C-sections are faster than vaginal births, when faced with time constraints, physicians may decide to put some patients through surgery -without a medical necessity for it-, reducing the midwives' workload. Within the agency discrimination framework, physicians will choose to practice an unnecessary surgery on patients with lower bargaining power. This study tests for the presence of agency by comparing single women and non-single women, assuming that single patients are -on average- more likely to be alone in the delivery room. In those cases, the physician will need less effort in convincing the patient to have a C-section. Indeed, the data shows that the gap in the probability of delivering by C-section between these two groups is statistically significant only for high ratios of patients-to-midwives. On the other hand, I find that married and low-risk patients are between 24% and 35% more likely of not attaining skin-to-skin contact with their newborn when the number of patients per midwife is high. This provides more evidence that, by performing more C-sections, physicians are avoiding some bad outcomes.

This paper contributes to several strands of literature. First, it adds to existing work on the effect of staff ratios on health outcomes. Previous studies find none or very small effects when using census discharge data (Evans and Kim, 2006; Cook et al., 2012), and a negative impact of crowding on health when focusing on patients in the Emergency Department (ED) (de Araujo et al., 2013). This difference across areas makes sense given the particular time constrains of patients in the ED. The maternity wards lay somewhere in between these two. However, there is no study looking at the effect of staff ratio in maternity wards using a casual approach. The one that comes closest to this is Balakrishnan and Soderstrom (2000), using data from 225,473 maternity admissions at 30 hospitals in the state of Washington. They identify crowded days using a percentile cut-off from the distribution of patients' admissions for each hospital-year combination and use the rate of C-sections as outcome. They find a positive and significant correlation between the two, but only for those pregnancies that are classified as at-risk of C-section. A shortcoming of this paper is that they cannot differentiate between scheduled and unscheduled patients in their data, raising concerns about causal relationships. It could be the case that days with more patients are those with more planned C-sections, without necessarily having any effect on patients' health outcomes. I contribute to this literature by causally estimating the effect of low staffing ratios on delivery method.

Second, there is a vast number of empirical studies that look at different causes for the exceedingly high levels of C-sections. Starting from the paper by Gruber and Owings (1996) where they use physician's income drop as a trigger for more C-sections, to other incentives like relative prices between C-sections and vaginal deliveries (Gruber et al., 1999; Alexander et al., 2013; Allin et al., 2015), defensive medicine (Keeler and Brodie, 1993; Lawthers et al., 1992; Currie and MacLeod, 2008; Dranove and Watanabe, 2009), and physician's scheduling convenience (Lefèvre, 2014).[11] I provide evidence of a new mechanism behind this phenomena: work overload provides incentives for physicians to induce C-sections.

Third, this study also relates to the literature that empirically tests possible mechanisms behind PID. Two recent papers use information asymmetry variations in the maternity ward set up. Grytten et al. (2011) compare expert and non-expert patients and conclude that a model of statistical discrimination (expert patients are better at communicating with the physician) explains their results better than one of agency discrimination (physician influences the diagnosis and treatment for non-expert patients). On the contrary, Johnson and Rehavi (2016) find evidence that physicians are more likely to exploit the information asymmetry when it is profitable. They do so by comparing physician patients with non-physician patients, in settings with and without financial incentives to perform C-sections. I add to this body of work by using a different approach to test for bargaining power: whether the mother is alone in the delivery room.

The remainder of this paper is organized as follows: Section 2 describes the clinical and institutional setting. Section 3 discusses the identification strategy followed and describes the data. Section 4 reports the results, and Section 5 concludes.

[11] For an extensive review of this literature see Allin et al. (2015).

1.2 Clinical and Institutional Setting

Maternity wards receive two types of patients: scheduled and unscheduled. The former includes patients admitted for an elective C-section and those who will be induced.[12] For patients following an elective C-sections the date of delivery is set in advance, and there is no possibility for changing delivery method (unless the mother goes into labor before). These pregnancies typically present some health condition that constitute a risk for the mother and/or the baby if delivered vaginally. Similarly, induced patients already know in advance the date they will be induced but, although they will attempt a vaginal delivery, the physician may still decide to change delivery method on the way if considered necessary.

The remaining patients -those attempting to follow the natural course of labor and vaginal delivery- are the main focus of this study. For these patients the process starts with frequent contractions and/or because they believe their water has broken (spontaneous onset of labor). Once the mother arrives to the hospital she is evaluated and if in active labor, she is admitted to the delivery room and assigned a gynecologist and a midwife. If everything goes as planned and the patient is able to have a vaginal delivery, the midwife will be the one helping her throughout the whole process. Nevertheless, during labor there are several medical conditions that can emerge and complicate a vaginal birth, putting in danger the health of the infant and/or the mother. Under these circumstances, the midwife and gynecologist may decide to recommend a C-section instead.

More importantly, the actual presence of some of these medical conditions depends heavily on the subjective opinion of the gynecologist.[13] This gray area -or asymmetry of information- on when a C-section is necessary gives the gynecologist more room to suggest the patient to have a surgery -even when not medically needed.

The maternity unit analyzed in this paper is part of one large teaching hospital in Italy. The staff working in the delivery room are paid a fixed salary, meaning they have no personal financial incentive to recommend any particular treatment. On the other hand, hospitals are reimbursed depending on a DRG (Diagnosed-related group) tariff system, which in general gives a higher reward for a C-section than a vaginal delivery.[14]

[12]Most inducements are performed on pregnancies that have past their due date and still haven't started labor.

[13]Two of these more 'subjective' conditions are dystocia (abnormally slow labor) and fetal distress

[14]For a deeper discussion on the Italian Health System please see Francese et al. (2014).

1.3 Empirical Methodology

1.3.1 A natural experiment

An ideal experiment to test for an effect of low-staffing in the maternity ward on patients' delivery method would imply assigning parturient women randomly between two different hospital types: a first one with already a large number of patients and a second type, identical to the first, but with few patients and hence ready to focus entirely on the coming patient. For obvious reasons this is not possible to implement in practice.

This paper focuses on patients who attempt vaginal delivery, and uses the exogenous variability in the number of patients and midwives present at admission to causally identify the impact of low staffing on delivery method. For the majority of births, the time of arrival is unknown to the hospital beforehand. In the same way, the level of capacity utilization of the maternity ward in a given point in time is unknown for future patients until they reach the hospital. For this sample of patients, their pre-admission probability of developing a complication and needing C-section is orthogonal to the level of crowding at the hospital.

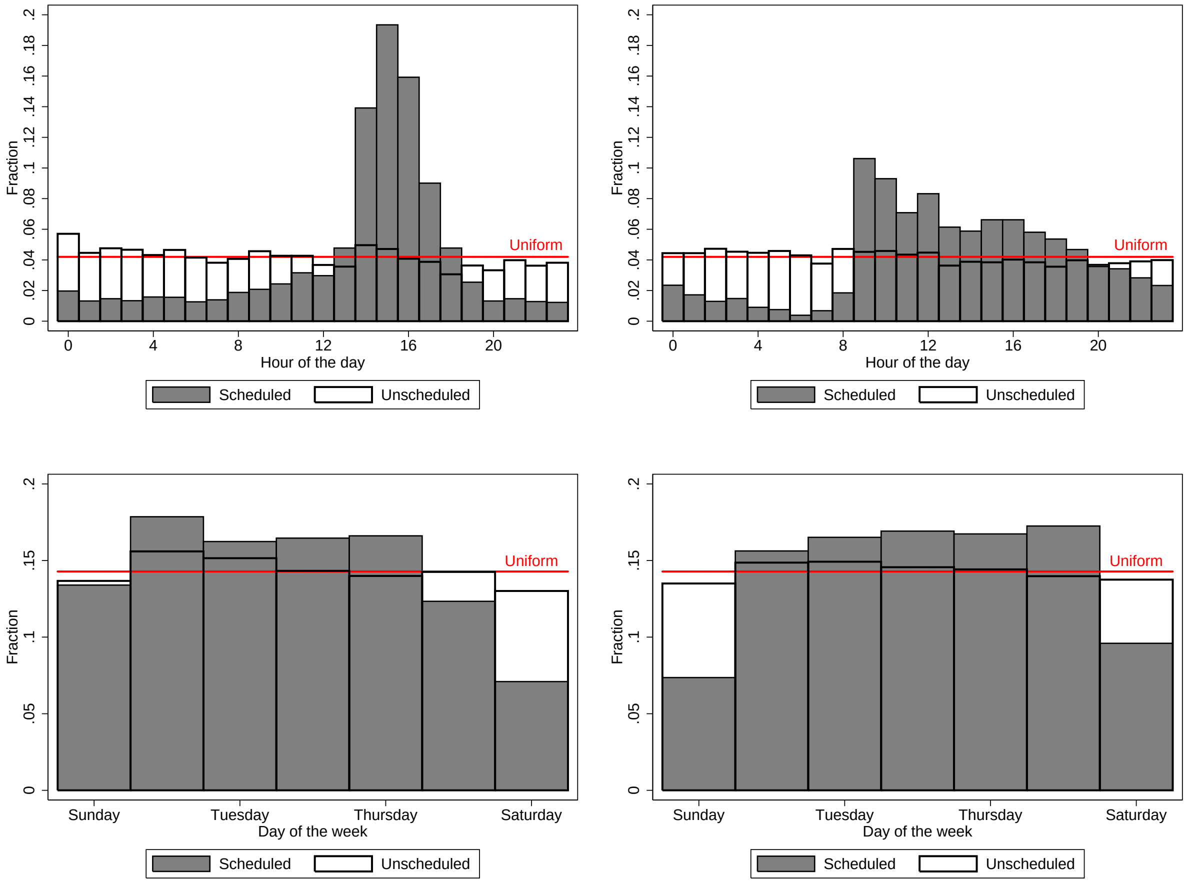
The study sample includes all births that, up to the point of arriving to the hospital, followed the "natural" course of pregnancy and labor. This means leaving out all scheduled deliveries where the physician decided, together with the patient, the date when the birth should take place. This type of patients are those who had an elective C-section or who were pharmaceutically induced to start labor.[15]

The left column of Figure 1.1 shows the distribution of admissions by hour of the day and day of the week. The right column does the same for births. Both are estimated for scheduled and unscheduled patients, for comparative purposes. We can immediately see that admissions of scheduled patients are concentrated in the afternoon, while births start at 9 a.m. and become less and less frequent as the day goes by. Instead, both admissions and births for unscheduled patients are very close to a uniform distribution across the day. When looking at the distribution by days of the week, again unscheduled patients are randomly distributed while scheduled patients are less common to be admitted on Saturdays, and less likely to have surgery on Sundays and Saturdays.

[15]For more evidence supporting the criteria for selecting the working sample see Appendix A.1.

Figure 1.1 – Distribution of admissions and births.

(a) By Time of Admission **(b)** By Time of Birth



1.3.2 Data

Previous studies looking at newborns' health tend to use anonymous birth certificates since they are publicly available for many countries and for long periods of time. Nevertheless these datasets commonly lack information on key variables needed for a rigorous study of staffing levels, namely the exact date and time of admission of patients (demand side) and the number of staff available (supply side), for each hospital.

This study utilizes data from the Maternity Department of the Azienda Ospedaliero Universitaria Careggi (AOUC) for the years 2011 through 2014. This is the biggest hospital in the Province of Florence with more than 3,000 deliveries per year. The primary databases used are two: (i) birth certificates[16]; and (ii) hospital admissions[17]. Birth certificates constitute a census of all births that took place in the hospital in this period. It contains information on mother characteristics (e.g. community of residence, education, civil status,

[16]Certificato di assistenza al parto (CEDAP).
[17]Scheda di Dimissione Ospedaliera (SDO).

age, previous deliveries, etc.), pregnancy characteristics (e.g. weeks of gestation, controls, assisted reproduction, etc.) and birth characteristics (e.g. time of birth, type of labor, attendant, place, weight of the baby etc.). The administrative hospital admission data provides information on the time of admission and time of discharge for each patient. Using unique mother-pregnancy identifiers, both databases can be merged together.

The aforementioned data on patients is complemented with information on the level of staff scheduled to be present at each month, day of the week and shift in the delivery room. Note that this is not the effective level of staff present at each point in time but the schedule that the personnel should follow. Anecdotal evidence suggests that deviations from planned levels are rare, even because the hospital calls in someone else when an employee misses her shift.

However, the richness of this dataset comes at a cost: because the information available corresponds only to one hospital in a four year period the sample size is relatively small. Furthermore, due to the path dependence of treatment in second and higher order births, this study focuses on first-time mothers. There were approximately 5,240 singleton births at this hospital in the sample period. From this, about 870 observations are plural births and/or delivered by urgent C-section which will not be taken into account in the analysis because of their particular characteristics and handling within the hospital. Further restricting the sample to non-induced planned-vaginal deliveries, the number of observations goes down to around 2,685. Finally, after dropping observations with missing time of admission, maternal age, education, birth order, weight and prenatal visits, the number of observations in the working sample is about 2,600. The models described below are fitted to this sample.

Table 1.1 summarizes the variables used in the analysis. The first column corresponds to the whole sample. Most of the patients who attempt a vaginal delivery succeeded. Only about 12% had an in-labor C-section. Patients are on average 31 years old, only 36% has a university degree, and 44% are single. There are few cases with bad outcomes: only 4.6% have a 5-minute APGAR score below 9, and about 5% are born prematurely or weighting less than 2,500 grams. Columns 2 and 3 report statistics for patients with a low and high ex-ante risk of C-section respectively. Columns 4 and 5 do the same by civil status. By construction, patients with high-risk are more likely to give birth by C-section, to use the neonatal intensive care unit, and to have an APGAR score below 9. They are also more likely to be single and less likely to have a university degree. Finally, single patients are less likely to have a university degree and more likely to delivery by C-section, although other outcomes are similar to the married subsample.

1.3.3 An exogenous measure of midwives' workload

A good measure of effective staff contains information on both the number of patients and personnel. For this case study I use the ratio between the

Table 1.1 – Descriptive statistics

	All	Low-risk	High-risk	Married	Single
Outcomes					
% vaginal birth	88.1	88.8	85.1	89.5	86.6
% in-labor C-section	11.9	11.2	14.9	10.5	13.4
Other interventions and health outcomes					
% operative birth	13.3	13.6	11.9	13.5	12.6
Average length-of-stay (hours)	76.0	75.3	79.0	76.7	75.9
% need of NICU	7.3	4.8	17.8	5.9	8.1
% lack of skin-to-skin contact	19.3	16.0	33.5	18.0	20.0
% non-exclusive breastfeeding	36.0	33.9	46.7	35.5	36.5
% APGAR score below 9	4.6	3.3	10.5	3.8	5.4
Mother's characteristics					
Average age	31.1	31.2	30.6	31.2	30.8
% with university degree	35.9	36.3	33.9	41.3	30.2
% single	44.2	43.5	47.0	0.0	100.0
Pregnancy's characteristics					
% born before 37 weeks of gestation	5.3	2.7	16.6	5.2	5.2
% with at least 1 ER visit	11.5	0.0	60.6	10.4	13.3
Newborn's characteristics					
% male	51.0	50.2	54.3	51.9	50.4
Average weight at birth	3,235	3,271	3,085	3,234	3,234
% low birthweight (<2,500 grams)	4.9	0.0	26.1	4.9	4.9
% high birthweight (>4,000 grams)	3.9	0.0	20.4	4.4	3.7
Observations	2,613	2,118	495	1,300	1,028

Statistics for main sample of unscheduled first-time mothers, from 2011-2014. High-risk are patients who, at admission, have a higher probability of needing a C-section. Those are defined as patients with newborns with extreme birthweight and patients with an emergency department visit during pregnancy. Low-risk are those without any of those characteristics.

number of patients and the number of midwives in the delivery room.[18] The richness of the data at hand allows to construct a very precise measure of the number of parturient women in the maternity ward at any point in time and to differentiate between those waiting to give birth and those in postpartum. But there are yet two decisions to be taken regarding the moment at which this ratio is calculated, and the type of patients to include in the numerator. On the former, because patients stay on average 7 hours in the delivery room between admission and birth, it is not obvious at what time to measure the level of staffing. The two most viable options are at the time of admission and at the time of delivery. The last one has the advantage of measuring staff when needed the most, meaning, when the mother needs help to give birth. The problem with this option is that, given that physicians can rush a delivery (e.g. by doing a C-section), the level of staffing at time of birth can be endogenously determined. On the other hand, even though the level of staffing at time of admission can be relatively less relevant, it is indeed an exogenous shock. For these reasons I will use the ratio of patients to midwives calculated at the time of admission of each patient.[19]

On the second issue, it is important to clarify which patients are included in this measure of staffing. The first option would be to include all patients (regardless of whether they are scheduled or induced). One could think that, because the time of the admitted patient is random, there is no risk of endogeneity here. Nevertheless, since the outcome of interest is the probability of C-section, counting elective C-sections in the measure of staffing would make it biased. Note that when there are more elective C-sections there are also more gynecologists ready to perform them. Incorporating elective C-sections in the numerator would not only include a demand side but also a change in the supply of physicians who can perform C-sections. Hence this study includes in the numerator all patients but those already scheduled to give birth by C-section.[20] Instead, the number of scheduled C-sections is included in the regression as control (see econometric specification below).

More specifically, the workload observed by patient i at admission time t is defined as

[18]One drawback of this measure is that it constraints the coefficient of interest due to the simultaneous variations in numerator and denominator. The fact that my preferred model specification uses fixed effects by shift and day-of-the-week means that all the variation used for the estimation comes solely from fluctuations in the numerator, alleviating this issue. Furthermore, Appendix A.2 repeats the main analysis using solely the number of patients as the covariate of interest, and results are qualitatively the same. In light of these results, in the main paper I will use the ratio of patients-to-midwives since it provides an advantage with regard to external validity (findings become less dependent on the size of the hospital studied).

[19]In the following section I perform several robustness check measuring staff levels at different points in time during a patients stay, and discuss the results.

[20]Note that this is not the same sample as the study sample because it also includes induced deliveries. Those are not at risk of contaminating the measure because they will still attempt a vaginal delivery, and will need a midwife to help them.

$$\mathrm{R}_{it} = \frac{PVB_{it}}{MW_{it}} \tag{1.1}$$

where PVB is the number of patients waiting to attempt a vaginal birth, and MW is the number of midwives scheduled to be present in the delivery room.

Table 1.2 shows the mean number of midwives and patients (with its standard deviation) in the delivery room by day of the week and shift of admission. The number of midwives is higher during the morning shift (5), and lower at nights and Sundays (3). On the other hand, the average number of patients is virtually the same across days of the week and shifts, with a slightly lower level on Sundays.[21]

Table 1.2 – Number of midwives and patients by day of the week and shift.

Day	Shift	Midwives (mean)	Midwives (sd)	Patients§ (mean)	Patients§ (sd)
Weekdays	Morning (7am - 1pm)	5	0	7.31	2.81
	Afternoon (1pm - 7pm	4	0	7.48	2.89
	Night (7pm - 7am)	3	0	7.32	2.86
Saturdays	Morning (7am - 1pm)	4	0	7.53	2.63
	Afternoon (1pm - 7pm	4	0	7.41	2.70
	Night (7pm - 7am)	3	0	7.26	2.71
Sundays	Morning (7am - 1pm)	3	0	7.09	2.73
	Afternoon (1pm - 7pm)	3	0	7.08	2.76
	Night (7pm - 7am)	3	0	6.94	2.68

§ Number of patients waiting who attempt to have a vaginal birth.

Table 1.3 shows the distribution of the ratio of patients to midwives for the whole sample and then disaggregated by shift of admission. The ratio is unimodal and slightly skewed to the right.[22] At the median, there are 2 patients for every midwife in the delivery room. The 25th and 75th percentiles are 30% (below) and 34% (above) the median, respectively. Note that shifts later in the day have higher values of the ratio, meaning, more crowding. Remember that the distribution of patients is rather uniform across the day, hence this upward shift in the ratio comes exclusively from a lower supply (less midwives present).[23] The bottom rows of the Table 1.3 show the cutoff values for the lowest and highest quintiles (and by construction for the three middle quintiles altogether). The lowest quintile will be considered a case with no crowding, with a mean of 1 patient per midwife. The middle quintiles have a mean ratio

[21] The difference with Sunday is due to the fact that there are less induced births.

[22] See Figure A.4 for a graphic representation of the density distribution of the ratio by shift.

[23] In Figure A.5 one can see how the average ratio of patients to midwives by hour of admission shows a discrete jump up with each change in shift due to one less midwife being present.

of 1.9, somewhat crowded. The highest ratio, with a mean of 3.2 patients per midwife, will be referred to as highly crowded or chaos.

Table 1.3 – Descriptive statistics for ratio of patients to midwives by shift of admission.

	All	Morning (7am - 1pm)	Afternoon (1pm - 7pm)	Night (7pm - 7am)
p1	0.60	0.40	0.50	0.67
p5	0.80	0.75	0.75	1.00
p25	1.40	1.20	1.50	1.67
p50	2.00	1.50	1.75	2.33
p75	2.67	2.00	2.33	3.00
p95	3.67	3.00	3.25	4.00
p99	4.50	4.00	4.50	4.67
mean	2.06	1.60	1.91	2.36
sd	0.86	0.70	0.74	0.88
<20th Percentile	1.33			
>80th Percentile	2.67			
Obs.	2,613	636	641	1,336

1.3.4 Econometric specification

The first part of the analysis estimates OLS regressions of a binary indicator for C-section on the treatment variable along with demographic and clinical controls. A simple reduced-form linear probability model of the following type is used:[24]

$$\mathrm{y}_{it} = \alpha + \beta\,\mathrm{R}_{it} + \theta\,\mathrm{X}_{it} + \gamma_t\,\mathrm{cs}_{it} + dowXshift_t + year_t + month_t + \epsilon_{it} \quad (1.2)$$

where y_{it} is a dummy variable indicating whether birth i admitted at time t had an in-labor C-section, and R_{it} is the ratio of patients-to-midwives observed at admission as explained above. X_{it} contains individual-level control variables of mother and pregnancy characteristics[25]. To further control for supply side

[24] A probit model was also estimated assuming a normal distribution of the error term and results are virtually the same (See Table A.6).

[25] These include: a dummy for whether the mother is above 34 years old, a dummy for whether the mother has a university degree, a dummy for whether this is her first pregnancy, a dummy for whether the infant is a male, a dummy for whether is a pre-term birth (below 37 weeks of gestation), a dummy for whether the baby is born with low weight (less than 2,500 grams), and a dummy for whether the mother had at least one emergency check up during pregnancy.

changes in physicians availability, I include the number of scheduled C-sections that took place while the indexed patient was in the delivery room (cs). Since most supply side changes in the maternity ward take place between shifts and days, in the most demanding specification I also add fixed effects for day-of-the-week (dow) times shift.[26] To control for seasonal and secular variation in outcomes, I also include monthly and yearly dummy variables. β is the coefficient of interest. As discussed above, if physicians are more likely to perform a C-section when the ratio of patients to midwives is high, then β should be positive.

Two models are estimated for the probability of delivering by C-section. First, I include the ratio of patients-to-midwives directly in the model. Because there can be non linear effects between staffing and delivery method, for the second model I split the sample in three categories based on the ratio of patients-to-midwives: low, medium, and high (or chaos). All those observations with a ratio below the 20^{th} percentile are in the first group. These are cases of no crowding, or very low ratio of patients to midwives. The second group includes those observations between the 20^{th} and 80^{th} percentiles, and are categorized as cases with some crowding. Finally, the last group consists of all those above the 80^{th} percentile. These are situations of very high ratios of patients to midwives. The cut offs for these groups are reported in the bottom of Table 1.3. In these models, the lowest quintile (low staffing) is considered the reference group.[27] Table A.4 shows the coefficients of a regression of each of the pre-treatment controls on the ratio of patients-to-midwives. The lack of statistical significance for all cases provides support to the exogeneity assumption of my measure of staffing. Furthermore, for the non-linear specification, Table A.5 shows that the mean of the pre-treatment characteristics are not statistically different across the three groups of staffing (low, medium and high). Again, this emphasizes the strength of the quasi-natural experiment.

The last part of the analysis aims at understanding the mechanisms through which physicians decide to recommend some patients to change delivery method. Two hypothesis are tested. First, it could be the case that high values of the ratio of patients to midwives results in less midwifery time available for each patient. Under this scenario, patients who were admitted with an already higher risk of C-section (and that need more care) will be the most affected. At higher ratios, the probability of C-section should rise faster for this group than for other patients -all else constant- due to their pre-treatment lower health. Patients with a higher risk are identified as those with extreme birthweight (below 2,500 grams or above 4,000 grams) or with at least one emergency visit to the hospital during pregnancy.

[26]This means that all the variation in this specification comes from within same day of the week and shift. For example, I would be comparing a mother who arrived on a Tuesday afternoon shift and finds many patients waiting with another woman arriving a different Tuesday afternoon but who observes few patients waiting.

[27]See Appendix A.3 for a more detailed discussion on model selection, where models of different polynomial degrees and categorical definitions of workload are tested.

The second hypothesis has to do with agency discrimination. When resources are constrained, e.g. high workload, physicians may consider it is optimal to shift some patients to the operative theater for a C-section. This would reduce the workload on midwives by reducing the number of patients waiting in the delivery room. Because patients are heterogenous, physicians will find it easier to offer this treatment to some patients more than others. This paper uses the patient's civil status as a proxy to whether she is alone in the delivery room.[28] The assumption here is that, on average, single women are more likely to be alone in the delivery room.[29] In those cases, the physician only needs to convince one person about the change in procedure -not to mention the patient is in labor and under stress, which makes harder to analyze the pros and cons of each alternative-.

To test whether physicians' treatment covaries with the patients' characteristics above mentioned, I estimate the following regression:

$$\mathrm{y}_{it} = \alpha+\beta_1\,\mathrm{R}_{it}+\beta_2\,\mathrm{R}_{it}\times\mathrm{D}_{it}+\beta_3\,\mathrm{D}_{it}+\theta\,\mathrm{X}_{it}+\gamma_t\,\mathrm{cs}_{it}+dowXshift_t+year_t+month_t+\epsilon_{it} \tag{1.3}$$

where D_{it} is either one of two variables: an indicator for whether the patient has a high-risk of C-section, or whether she is single. The remaining variables are defined as in Eq. (1.2), adding civil status as a control. I expect high-risk and single patients to be more affected by a high ratio of patients, hence, a positive β_2 in both cases.

1.4 Results

Table 1.4 presents the results of Eq. (1.2). Starting from a regression with only the covariate of interest and fixed effects for year, month and day of the week in the first column, each remaining column sequentially adds more controls. The second column adds controls for mother and pregnancy characteristics, the third adds the number of scheduled C-sections taking place during patient's labor, the fourth column includes hour of admission fixed effects, while the last one instead uses shift-of-admission interacted with day-of-the-week fixed effects. This last model is the preferred one since it accounts for possible supply changes in the ward that occur between shifts and days. To save space, only the coefficients of treatment are included, but results for other covariates are comparable to previous studies.[30] The numbers in parentheses in the table are standard errors. The average value of each dependent variable is included at the bottom of each panel to help understand whether coefficients are economically important. For all remaining estimations in this paper I will use the specification model in column (5).

[28] This variable is constructed only with married and single women. For the sake of clarity, all women outside these two categories (divorced, separated and widows) are not considered.

[29] For a single woman in Tuscany, the odds of being alone in the delivery room are 1.25 times larger than the odds for a married woman being alone (**(alias?)**).

[30] Full regressions are available upon request.

Table 1.4 – Effect of effective staffing on the Probability of C-section

	(1)	(2)	(3)	(4)	(5)
Panel (A)					
Ratio patients	0.0019	0.0038	0.0046	0.0007	0.0004
to midwives	(0.0074)	(0.0074)	(0.0074)	(0.0083)	(0.0085)
Panel (B)					
20-80th Percentile	0.0432***	0.0441***	0.0452***	0.0410***	0.0408***
	(0.0149)	(0.0148)	(0.0148)	(0.0154)	(0.0156)
>80th Percentile	0.0193	0.0238	0.0261	0.0187	0.0197
	(0.0196)	(0.0195)	(0.0196)	(0.0210)	(0.0213)
Observations	2,613	2,613	2,613	2,613	2,613
Mean dep.	0.119	0.119	0.119	0.119	0.119
Time FE	✓	✓	✓	✓	✓
Controls		✓	✓	✓	✓
Other patients			✓	✓	✓
Hour FE				✓	
Shift*dayofweek FE					✓

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$
Reported coefficients are average marginal effects.

Panel (A) of the table reports results for the ratio of patients-to-midwives as a continuous variable, and Panel (B) reports results using a dummy variables for different levels of workload in order to test for non-linearities. Notice that coefficients across columns (models) only change in the third decimal. This is a good sign of exogeneity of the ratio of patients-to-midwives. Although the coefficient for the linear specification is not statistically significant, in the second panel the probability of having a C-section is about 4 p.p. (34%) higher for those who face a ratio of patients-to-midwives in the middle of the distribution compared to those in the reference group. For those patients arriving when the ratio of patients-to-midwives is very high (last quintile), there is no statistically significant effect on the probability of C-section. This may be due to some capacity constraints on the operative theater when workload is at its highest levels.

This effect would imply a 5.7% (or 1.2 p.p.) rise in total C-sections (scheduled and unscheduled), which is economically important and reasonable when compared with previous studies looking at all C-sections and changes in monetary compensation. Allin et al. (2015) find that doubling the compensation for a C-section relative to a vaginal delivery increases the likelihood that a physician opts for the former by just more than 5 p.p., all else equal. Gruber et al. (1999) suggests that cesarean delivery rates would rise by 3.9% in response to each $100 increase in the compensation received for a C-section, all else equal.

Table A.7 presents results of the effect according to whether the patient arrived in a weekday or weekend, and by shift of admission. The estimations are very imprecise due to the few number of observations in each cell, and render all differences insignificant. Nevertheless point estimates are slightly higher in weekends, as well as for admissions during the morning shift. Table A.8 shows results for a robustness check where I measure effective staff level at different points in time between a patient's admission and delivery. The effect of congestion disappears the further away from admission it is measured, which can be a result of the endogeneity issue mentioned before: physicians can adjust the timing of births. Finally, Table A.9 presents results for a placebo test where workload is measured 24 hours after admission (instead at admission as before). As expected, for all different specifications, the placebo is always statistically and clinically insignificant.

1.4.1 How do physicians choose which patients to send to the operative theater?

This part of the study digs deeper into the mechanisms behind the effect of staffing on the rate of C-sections. As mentioned before, two hypothesis are tested. First, low-staffing means there is less midwifery-time available for each patient, which may result in more patients needing C-section due to the lack of proper care. This effect should be higher for those patients who were admitted with an already higher risk of C-section. Secondly, physicians and midwives may actively decide to perform a C-section on some patients in moments of low-staffing in order to reduce the number of patients in the delivery room. In this case I expect patients with lower bargaining power -which I proxy by civil status- receiving relatively more unnecessary treatment.

Table 1.5 reports the average marginal effects obtained for each group from estimating Eq. (1.3). As expected, a higher number of patients per midwife rises the probability of C-section more for single patients but not for married ones. Points estimates suggest that high-risk patients are more affected by workload than low-risk patients, although the only statistically significant coefficient is for this group. However estimates are very imprecise.

Another way to look at it is by comparing the average marginal effects of being high-risk and single, across the different levels of the ratio of patients-to-midwives. This can be seen in Figures 1.2a and 1.2b respectively. Note that the effect of staffing, in both cases, is not statistically significant when the ratio is low. For the comparison based on ex-ante risk, the point estimate for the difference in the probability of C-section between the two groups gets higher with workload -albeit not statistically significant-. This is reasonable since the marginal benefit from midwives' attention is higher for high-risk patients.

Instead, for the case of married vs. single mothers, the difference is statistically significant only for those in the middle of the distribution, but goes down again

Table 1.5 – Effect of effective staffing on the Probability of C-section

	Low-risk	High-risk	Married	Single
20-80th Percentile	0.0330*	0.0587	0.0270	0.0516**
	(0.0175)	(0.0368)	(0.0202)	(0.0247)
>80th Percentile	0.0023	0.0755	0.0125	0.0182
	(0.0229)	(0.0557)	(0.0265)	(0.0332)
Observations	2,118	495	1,300	1,028
Mean	0.11	0.15	0.11	0.13

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$. Reported coefficients are average marginal effects from a regression of the probability of C-section on the interaction of treatment, a variable for being high risk and a variable for being single. The number of observations when using marital status is slightly smaller because the variable is missing for 11% of the working sample.

when workload is high. At high levels of workload it is more likely that capacity constraints in the operative theater emerge as well. These "extra" C-sections only based on midwives' workload and not due to patients' health-status should go down during the busiest times.

1.4.2 The effect on other interventions and morbidity outcomes

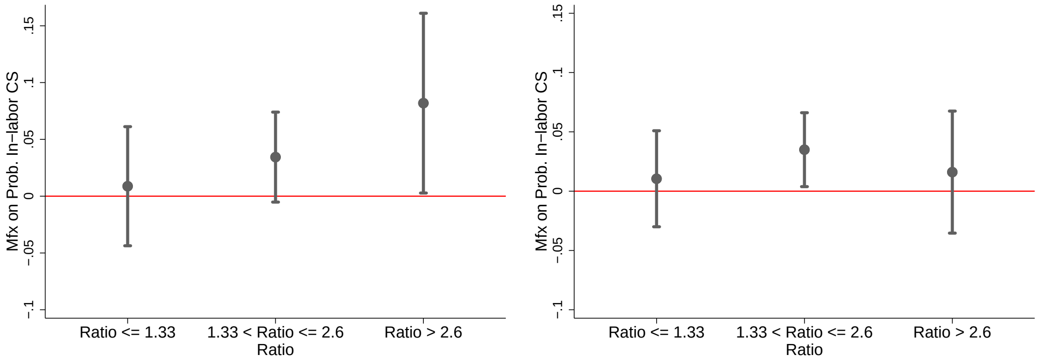
The estimates above demonstrate that, when the ratio of patients-to-midwives is high, physicians send some patients to the operative theater to have a C-section. These patients are typically those with a higher-risk of needing a C-section, or single women. The question that emerges then is: are physicians using their high bargaining power to transfer some patients so that midwives can provide better care for the remaining patients? In order to test this, I estimate Eq. (1.3) again but now the outcome variable is one of the five indicators of morbidity and interventions mentioned before. If a high ratio lowers the quality of care, then those type of patients who are not likely to be sent to the operative theater would be the ones more affected by it.

In the economics literature the most commonly studied health outcomes for births are: weight, fetal mortality and maternal mortality. Nevertheless both maternal and fetal deaths are extremely rare events (4 per 100,000 births and 2.7 per 1,000 births respectively for Italy). In the case of weight-at-birth, because treatment here is defined at the moment of admission to the hospital, it is considered a pre-defined outcome (not affected by treatment).[31]

[31] In fact weight at birth is one of the variables used to assess the balancing of the sample between treatment and control groups.

Figure 1.2 – Difference in the effect of staffing on the probability of C-section by type of patient

(a) By pre-admission risk **(b)** By civil status



Note: Dots are the average marginal effect of whether the patient is high-risk (a) or single (b). Bars are 90% confidence intervals.

The restricted-use version of the birth certificates at hand contain, however, some other measures of health and registers of medical interventions that are associated with health outcomes. The measures that occur in at least 1% of births are: having an operative birth[32], length-of-stay after birth (LOS), whether the newborn was transferred to a neonatal intensive care unit (NICU), no skin-to-skin contact, lack of exclusive breastfeeding, and whether the newborn had an APGAR score[33] below 9. A higher probability of needing NICU, having an operative birth[34] or a longer time in the hospital during crowded times can be signals of lower quality. Similarly, if human resources are scarce, physicians may decide to skip some steps of the service considered important but not essential. For example, they may decide that helping the newly mother achieve skin-to-skin contact with her newborn is not as important as helping another woman in labor to deliver. The same reasoning applies for not giving exclusive breast-feeding.

While it is clear why a higher probability of going to NICU, having a low APGAR score, or staying longer in the hospital are not desirable, there are also compelling arguments regarding the importance of the remaining set of outcomes. In a systematic review, Ip et al. (2007) finds that breastfeeding is associated with both decreased risk of many early-life diseases and conditions as well as with health benefits to women[35]. At the same time, skin-to-skin

[32] Operative vaginal delivery refers to a delivery in which the physician uses forceps or a vacuum device to assist the mother in transitioning the fetus to extra-uterine life.

[33] The Apgar score is a method used to quickly summarize the health of newborn children. The Apgar scale is determined by evaluating the newborn baby on five simple criteria on a scale from zero to two, then summing up the five values thus obtained. The resulting Apgar score ranges from zero to 10.

[34] A higher likelihood for operative birth has been linked to scarce or absent midwifery care and the presence of obstetrician or physicians instead (Hatem et al. (2008)).

[35] "Breastfeeding and Maternal and Infant Health Outcomes in Developed Countries", AHRQ Publication No. 07-E007, April 2007.

contact has been shown to increase the probability and length of exclusive breastfeeding (Moore et al., 2007), as well as to substantially reduce neonatal mortality amongst preterm babies in hospital(Lawn et al., 2010). In the case of operative births, even though it is still widely used, this delivery method is becoming less popular due to some evidence showing it increases maternal morbidity and can cause significant fetal morbidity (Ali and Norwitz, 2009; Murphy et al., 2011; Towner et al., 1999).

Table 1.6 displays the average marginal effects for each of the four groups of women (high and low risk, married and single), and for the five outcomes above mentioned. Estimates are quite imprecise given the small sample size and the rarity of these morbidities. However, there is a statistically significant, large and positive effect of the high ratios of patients-to-midwives on the probability of not achieving skin-to-skin contact with the infant. Furthermore, this effect is only present for married patients, who are not more likely to get surgery when workload rises. These patients are between 24% and 35% more likely to not attain skin-to-skin contact with their newborn when the number of patients per midwife is higher. This provides further evidence of the hypothesis that physicians send some patients to the operative theater in order to avoid other negative health outcomes.

1.4.3 Other possible channels?

Beyond the mechanisms mentioned in the previous section, there are, at least, two more channels that can explain the rise in C-sections along with the rise of the ratio of patients-to-midwives. The first and most obvious option is that patients who are admitted in low and high staffing times are different. Nevertheless, all tests performed in this study and previous research support the idea that, for those patients attempting a vaginal delivery, their time of arrival to the hospital is randomly distributed across the day and week.

The other possible explanation is that those type of patients who get these 'extra' C-sections actually have a preference for this delivery method. However, because the focus is exclusively on in-labor C-sections, the above estimates correspond to women who have already agreed on attempting a 'natural' vaginal delivery. Hence the effect is more likely to arise from decisions made in the delivery room regarding when to stop labor and change treatment, than from maternal preferences for C-sections. Nevertheless, because data comes from a public hospital, patients may be denied an elective C-section -even when preferred- if there is no medical reason for it. Hence it is not possible to totally rule out that some demographic groups may be more inclined towards having a C-section and physicians internalize this when deciding which patient to send to surgery.

Table 1.6 – Effect of effective staffing on other health outcomes

	Low-risk						High-risk					
	Op. birth	LOS⋆	NICU§	No s2s‡	NEB†	Apgar<9	Op. birth	LOS⋆	NICU§	No s2s‡	NEB†	Apgar<9
20-80th Percentile	-0.0164	-0.0246	-0.0134	0.0298	-0.0325	-0.0096	-0.0616	-0.0085	-0.0013	0.0133	-0.0392	0.0070
	(0.0194)	(0.0184)	(0.0123)	(0.0216)	(0.0296)	(0.0104)	(0.0389)	(0.0414)	(0.0304)	(0.0497)	(0.0629)	(0.0298)
>80th Percentile	0.0257	-0.0380	-0.0064	0.0426	-0.0179	-0.0011	-0.0213	-0.0168	-0.0357*	0.0329	-0.0592	-0.0075
	(0.0264)	(0.0234)	(0.0160)	(0.0287)	(0.0384)	(0.0144)	(0.0270)	(0.0274)	(0.0195)	(0.0310)	(0.0422)	(0.0181)
Observations	2,613	2,521	2,609	2,297	2,044	2,613	2,613	2,521	2,609	2,297	2,044	2,613
Mean dep.	0.133	4.274	0.0728	0.193	0.360	0.0463	0.133	4.274	0.0728	0.193	0.360	0.0463

	Married						Single					
	Op. birth	LOS⋆	NICU§	No s2s‡	NEB†	Apgar<9	Op. birth	LOS⋆	NICU§	No s2s‡	NEB†	Apgar<9
20-80th Percentile	-0.0339	-0.0061	0.0124	0.0449*	-0.0183	-0.0033	-0.0213	-0.0168	-0.0357*	0.0329	-0.0592	-0.0075
	(0.0241)	(0.0222)	(0.0140)	(0.0265)	(0.0367)	(0.0124)	(0.0270)	(0.0274)	(0.0195)	(0.0310)	(0.0422)	(0.0181)
>80th Percentile	0.0096	-0.0194	0.0111	0.0699**	-0.0517	0.0117	-0.0119	-0.0159	-0.0086	0.0416	0.0210	-0.0142
	(0.0329)	(0.0291)	(0.0171)	(0.0345)	(0.0472)	(0.0178)	(0.0359)	(0.0342)	(0.0272)	(0.0432)	(0.0559)	(0.0235)
Observations	2,328	2,264	2,324	2,057	1,826	2,328	2,328	2,264	2,324	2,057	1,826	2,328
Mean dep.	0.133	4.274	0.0728	0.193	0.360	0.0463	0.133	4.274	0.0728	0.193	0.360	0.0463

Reported coefficients are average marginal effects. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$
⋆LOS: Length-of-stay after birth (in log-hours); §NICU: Neonatal Intensive Care Unit; ‡No s2s: No skin-to-skin contact; †NEB: Non-exclusive breastfeeding.

1.4.4 Can these 'extra' C-sections be avoided?

Results above suggest that physicians do more surgeries when staffing is low. First-time mothers facing a ratio of patients-to-midwives between 1.33 and 2.66 are 4 p.p. (or 34%) more likely to have an in-labor C-section. A policy to eliminate overcrowding from maternity wards would have a very significant effect on the already high levels of C-sections seen in Italy. How to do that is not clear.

Considering only the hospital used in the analysis, in the absence of crowding, the "extra costs" for the public health system are of about €17,700 a year.[36] This is not enough to hire the necessary number of midwives to avoid low-staffing situations. However, one should keep in mind that this analysis is not complete, since one should include other costs like the drop in skin-to-skin contact when staffing is low, or the other non-financial costs of C-sections mentioned in the introduction of this study.

Another possible policy is to concentrate maternity wards in fewer but bigger units and benefit from the economies of scale emerging. The larger the population a hospital serves, the lower the coefficient of variation of demand, and hence the higher the occupancy rate (Long and Feldstein, 1967). For the hospital under study this may not really be a suitable alternative since it is already a large maternity ward and the only one in the city.

1.5 Conclusions

In this paper I use a natural experiment set up and detailed data on births to estimate the impact of staffing on physician's treatment decisions. More specifically, I investigate whether different levels of midwifery effective staffing (patients-to-midwives) influence the probability that a patient will be sent to have a cesarean section. The contribution is threefold. First, it proposes an innovative empirical approach that allows me to estimate physician's responses to exogenous shocks to effective staffing. Second, it provides suggestive evidence that physicians do not choose at random which patients to over-treat, but may instead exploit their bargaining power. Lastly, it brings to light yet another cause for the high C-section rates we see today: low effective staffing.

Focusing exclusively on patients attempting labor and vaginal delivery, this study finds that first-time mothers who -at admission- face a ratio of patients-to-midwives higher than 1.3 are about 34% (or 4 p.p.) more likely to change delivery method. There are two type of patients who are more affected by

[36] Back of the envelope calculations suggest that there are about 86 "extra" C-sections in the 4 years in the sample due to crowding. According to the prices on acute interventions published by the Italian Ministry of Health, a vaginal delivery without complication is rated at €1,272, while a C-section costs €2,092. Hence the difference (€820time the number of extra C-sections (107) divided by the number of years (4) gives €17,700.

this. First, patients who upon admission have an already higher risk of C-section are more likely to develop complications due to limited care when few midwives are available. Secondly, single women, due to their lower bargaining power. I provide evidence that physicians may decide to induce some patients towards having a C-section in order to speed up the delivery and release the pressure on midwives in the delivery room. In summary, the evidence provided here suggests that physicians' way to deal with an exogenous shock in demand (patients) is to induce some patients towards an intervention that is faster, maximizing the aggregate health in the maternity ward.

My estimates imply that the total number of C-sections for first-time mothers could be reduced by about 5.7% (1.2 p.p.) if situations of low-staffing are avoided. This would be a very important achievement given the already overly high rates of C-sections observed in developed countries. Nevertheless, it is not clear that public healthcare systems can quickly afford to tackle this issue.

Chapter 2

Forgetting-by-not-doing: The case of surgeons and cesarean sections

2.1 Introduction

There is a large medical literature that documents an association between higher hospital or surgeon procedure volume and lower mortality rates, for a wide variety of different procedures, time periods, and locations (Luft et al., 1979; Halm et al., 2002; Birkmeyer et al., 2002). The two main competing hypothesis the literature offers for this correlation are the 'learning-by-doing' (or 'practice-makes-perfect') and the 'selective referrals' effects (Luft et al., 1987). Under 'learning-by-doing', increased experience leads to improvement in skills which in turn results in better outcomes. 'Selective referral', instead, occurs when providers with higher quality attract greater demand and henceforth have a greater volume of patients. The importance of identifying which one is driving the correlation between volume and outcome stems from the fact that they have opposite policy implications. If volume causes outcome, as learning-by-doing suggests, then the concentration of procedures in fewer and bigger providers would bring better outcomes. However, if causality runs from outcome to volume, then those benefits are not present anymore, and concentration would only lead to reduced competition between providers.

This paper aims at causally identifying whether learning-by-doing is present at the individual level in the healthcare sector, specifically for surgeons performing cesarean sections (C-sections). In particular, I look at whether surgeon's recent procedure volume (measured by the number of C-section surgeries performed in the previous 30 days) affects performance (measure by patient outcomes). In order to establish a causal relationship, I benefit from the fact that, due to state regulation, pregnant women in Italy cannot choose the gynecologist that will help them deliver their baby within the public system.[1] This institutional feature creates a scenario where selective referral is not possible.

I make use of a census of birth certificates from a large public hospital in Italy for the period 2011-2014 that contains surgeon identifier for each operation. To address possible concerns that physicians with more or less skills might treat patients with a higher or lower risk (selective allocation), I use a fixed effect model and rely on changes in volume within surgeon for the estimation. I find strong evidence of learning-by-doing for C-section surgeons: performing a single additional procedure in the previous 30 days lowers the newborn's probability of having a low Apgar score (below 9)[2] and of being passed to a neonatal intensive care unit (NICU), both in a clinically meaningful and statistically significant way. This effect is only present for emergent C-sections (not for elective C-sections), meaning, cases in which the surgeon has to make crucial decisions against the clock.

Cesarean sections are an attractive procedure to analyze the presence of surgeon's 'learning-by-doing' hypothesis. Unlike other highly studied procedures that are performed by a team of surgeons, C-sections are executed by only one surgeon, allowing for better estimates of the individual surgeon's learning curve. In addition, for many developed countries, C-sections have become the most common surgical procedure.[3] Furthermore, a recent wave of closures of maternity services in various countries (e.g. US, Canada, UK, Japan, France, the Netherlands, and others)[4] makes the discussion on volume-outcome effects all the more relevant. To the best of my knowledge, this is the first paper that tries to obtain causal estimates of learning-by-doing for the case of cesarean sections.

The remainder of this paper is organized as follows: Section 2 provides a review of the relevant literature, while Section 3 describes the clinical and institutional setting. Section 4 discusses the identification strategy followed and describes the data. Section 5 reports the results, and Section 6 concludes.

2.2 Literature Review

In the recently published Encyclopedia of Health Economics, Vivian Ho states that *"...there are hundreds of papers in the medical literature finding an association between higher hospital or surgeon procedure volume and lower mortality*

[1]They can choose provider if they pay for it, but there are very few of these cases given the high reputation of the public system.

[2]The Apgar score is a method used to quickly summarize the health of newborn children. The Apgar scale is determined by evaluating the newborn baby on five simple criteria on a scale from zero to two, then summing up the five values thus obtained. The resulting Apgar score ranges from zero to 10.

[3]In the US, in 2011 there were almost 1.3 million C-sections while only 560,500 CABGs (Pfuntner et al., 2013).

[4]Anecdotal evidence: Healthy Debate-Canada, Womens Enews-US, The Guardian-UK.

rates. However, most rigorous econometric analyses of health care data have been unable to formally identify learning by doing."(Ho, 2014)

Studies on learning-by-doing have focused almost exclusively on two procedures for heart disease -coronary artery bypass graft (CABG) and percutaneous transluminal coronary angioplasty (PCTA). They typically use lagged or cumulative volume at hospital levels as the covariates of interest, and find no support for the learning-by-doing hypothesis (Gaynor et al., 2005; Ho, 2002; Sfekas, 2009). Instead, there is some evidence that the volume-outcome relationship is due to increasing returns to scale (Gaynor et al., 2005).[5]

The literature testing for volume-outcome effects using physician level data is much more limited (Ho, 2014), and finds no clear consensus. On the one hand, Huesch (2009) and Contreras et al. (2011) fail to find any association between cumulative surgeon procedure volume and patient's health. Huesch (2009) looks at 57 "new cardiac surgeons", a group where one would expect learning effects to be the greatest. The author uses a model with physician fixed effects and instruments volume with a choice model to mitigate potential issues of selective referral -although he does not reject exogeneity of volume for the physicians in the sample. In a different paper, Contreras et al. (2011) use a longitudinal census for a specific eye surgery (LASIK) in one clinic in Colombia. Their set up has three main advantages: (i) LASIK surgeries have precise measures of presurgical condition and postsurgical outcomes (eyesight); (ii) it is free of selective referral, since most patients are randomly assigned to doctors in order to equalize workload; and (iii) unlike many surgeries (e.g. CABG, PCTA), LASIK surgery is performed by one surgeon, and therefore suitable for capturing surgeon's individual learning curve. They use a model with surgeon fixed effects and use individual cumulative surgeries to test for learning.

On the other hand, Ramanarayanan (2008) and Huckman and Pisano (2006) find evidence of strong learning-by-doing effects at the physician level when using a measure of recent experience as their covariate of interest. This type of volume-outcome relationship has been called learning-from-recent-experience by Huesch and Sakakibara (2009). Ramanarayanan (2008) uses a longitudinal dataset on CABGs that took place in Florida in the period 1998-2006, and found that individual physician volume does have an effect on patient outcomes. To alleviate issues of endogeneity, the author uses the departure of a surgeon as an exogenous shock on the volume of the remaining physicians. Instead, Huckman and Pisano (2006) do not discuss potential bias to a great extent, and they confine themselves to using surgeon and hospital risk-adjusted mortality as quality controls. They also focus on CABG cases -although their data comes from Pennsylvania for 1994 and 1995- and find that the mortality rate of patients decreases significantly with increases in the surgeon's experience in the previous calendar quarter. Both these papers find that the effect of experience is mainly driven by hospital specific experience, and is only partially

[5]See Huesch and Sakakibara (2009) for a schematic revision of the literature on the volume-outcome relationship and possible mechanisms behind it.

(if at all) portable across hospitals.

The main strength of this paper lies with the absence of selective referral in the data. Although previous studies on CABG have gone a great length to find exogenous instruments to mitigate this issue, identification becomes considerably cleaner and straightforward in a context where selective referral is simply not possible. The one study using data with the same feature is Contreras et al. (2011). However, the present paper still holds the comparative advantage of looking at a procedure that is far more common than the eye surgery operation they focus on.

2.3 Background

2.3.1 The performance and organization of cesarean sections

A Cesarean section (C-section) is a major surgical procedure in which a fetus is delivered through an incision in the mother's abdomen and uterus (American College of Obstetricians and Gynecologists, 2010). The procedure typically takes 45 minutes to an hour, and most mothers and babies stay in the hospital for two to three days. C-sections are now the most common surgery in several developed nations, making them the focus of many policy discussions.[6][7]

Based on their urgency degree, C-sections can be classified in three groups: elective, in-labor (or emergent) and urgent.[8] The first group includes all scheduled C-sections, in which physician and mother have agreed in advance to not have a vaginal delivery -in general on the basis of an obstetrical or medical indication- but there is no maternal or fetal compromise. In-labor C-sections occur when, a patient who is already experiencing labor and attempting to have a vaginal delivery, develops some complications that put in danger the health of the infant and/or the mother and thus the physician recommends to change delivery method towards surgery. Finally, in urgent C-sections there is some maternal or fetal compromise which is not immediately life-threatening.

[6] In US hospitals, "...Cesarean section was the most common major operating room procedure performed in 2011 [...] and the hospitalization rate for stays with Cesarean section increased 39 percent since 1997." (Pfuntner et al., 2013).

[7] " Caesars legions", The Economist, August 2015.

[8] This classification is very close to the one proposed by Lucas et al. (2000), which encompasses 4 categories because it differentiates between scheduled and elective C-sections. Lucas' is the only classification accepted officially by the Royal College of Obstetricians and Gynaecologists and the National Institute for Health and Clinical Excellence.

2.3.2 The Italian health care system and C-sections

Italian health care is a universal, public-private insurance system. The public part is the national health service- Sistema Sanitario Nazionale (SSN)-, which is administered on a regional basis. According to the World Health Organization, the Italian system provides the second best overall health care in the world -the first one being France.(Organization, 2000)

Under this system, a pregnant woman cannot choose the physician that will follow her and delivery her baby, unless they pay. Furthermore, given the well functioning of the system, the grand majority of people choose to use the public service and do not choose.[9] This institutional feature limits the risk of selective referral, where institutions or surgeons with better performance will attract higher volumes of patients -a common endogeneity issue in studies of learning-by-doing.

The data used in this study comes from the Maternity Department of the Azienda Ospedaliero Universitaria Careggi (AOUC), a public-teaching hospital located in Florence (Italy). In this hospital patients are assigned to the physician in shift at the time of admission. However, more complex cases are handled by the most experienced person in that field -selective allocation-, which may bias my estimates towards zero. In order to mitigate these concerns, I perform several robustness checks that are described in the following section.

2.4 Empirical methodology

2.4.1 Empirical model

The main question addressed in this paper is whether there is learning-by-doing in cesarean section surgeons. I test this by looking at whether surgeon's recent experience (e_{st}) has an impact on next surgery's outcome. Thus I estimate a reduced-form model of the following type:

$$\mathrm{y}_{ist} = \alpha + \beta\,\mathrm{e}_{st} + \delta\,\mathrm{d}_{st} + \mathbf{x}_{it}^{'}\,\theta + \phi_t + \eta_s + \epsilon_{ist} \tag{2.1}$$

where y_{ist} is a health indicator for patient i whose procedure was performed by surgeon s at time t. Surgeon's recent experience is defined as the number of C-sections performed in the 30 days leading up to and including the procedure on the patient surgeon s operated on just before operating on patient i. d is a control for the number of days since the prior cesarean section surgeon

[9]For the data in hand, I can observe which patients have paid, and they constitute a few dozen cases that are dropped from the study sample.

s performed.[10] $\mathbf{x}_{it}$ contains individual-level control variables for mother and pregnancy characteristics.[11] Time fixed effects ϕ include year, month and day of the week.

Individual surgeon fixed effects (η_s) are included to mitigate concerns that the captured relationship between outcomes and recent experience is driven by composition effects. For example, high quality surgeons may choose to operate fewer patients but with higher risk of developing complications. Surgeon fixed effects ensures that the recent experience parameter in 2.1 is identified from changes in volume *within* surgeon. As discussed above, if physicians skills improve with repetition, then β should be negative: since outcomes are defined as adverse, a higher recent volume of surgeries would help (partially) avoid the lose of skills. On the contrary, a coefficient close to zero would imply that there is full depreciation and recent experience does not affect current outcome.

2.4.2 Robustness checks

Even if physician fixed effects help alleviate issues of selection of patients based on physician's skills, there could still be problems of endogeneity if there exist some sort of dynamic matching. For instance, a physician aware of depreciating skills may handpick healthier patients to operate on after a period of low activity, and my estimates on the impact of recent experience on patients' health would suffer from a downward bias.[12] To mitigate this concerns, I perform a series of robustness checks following Hockenberry and Helmchen (2014).

First, I estimate a separate coefficient for different types of C-sections, depending on their emergency status. Since in-labor C-sections are classified as emergent cases, where patients needing to be taken care of as soon as possible, there is little room for patient selection. Moreover, given the unexpected nature of these cases, surgeons need to make fast decisions under pressure and skill depreciation should be of particular relevance. These patients are already in labor, which adds one additional layer of complication: one important factor for the success of these surgeries is the timing of the cut in relation to the contractions.

In a second robustness check, I use a restricted sample with surgeons performing a minimum of 20 C-sections per year. This will help mitigate the possibility

[10]This measure is more precise than fixed calendar months as it responds instantaneously to any changes in the recent experience profile.

[11]These include: a quadratic term for mother's age, a dummy for whether the mother has a university degree, a dummy for whether this is her first pregnancy, a dummy for whether the infant is a male, a dummy for whether is a pre-term birth (below 37 weeks of gestation), a dummy for whether the baby is born with low weight (less than 2,500 grams), and a dummy for whether the mother had at least one emergency check up during pregnancy.

[12]The opposite case, were less active surgeons choose patients with worse health, is less likely to occur. Under this scenario, my estimates would be an upper bound of the true effect.

that my estimates of the effect of recent experience on outcomes are capturing some systematic unobservable quality differences between high and low volume surgeons and patients' health.

Finally, I test whether recent experience is correlated with pre-treatment characteristics that might be associated with patient's health.

2.4.3 Data

This study utilizes birth certificates[13] from the Maternity Department of the Azienda Ospedaliero Universitaria Careggi (AOUC) for the years 2011 through 2014. This is the biggest hospital in the Province of Florence with more than 3,000 deliveries per year. Birth certificates constitute a census of all births that took place in the hospital in this period. It contains information on mother characteristics (e.g. community of residence, education, civil status, age, previous deliveries, etc.), pregnancy characteristics (e.g. weeks of gestation, controls, assisted reproduction, etc.), birth characteristics (e.g. time of birth, type of labour, attendant, place, etc.) and indicators on newborns' health (e.g. weight, height, apgar score, death, etc.). This information is complemented with surgeon's ID.[14]

The richness of this dataset comes at a cost: because the information available corresponds to just one hospital in a four year period, the sample size is relatively small. There were approximately 12,343 newborns during the period under study, from which 4,413 (35%) are C-sections -the rest are vaginal deliveries-. Almost half of these C-sections are planned in advance between the physician and the patient (elective C-sections). The remaining half is more or less equally constituted by in-labour C-sections -pregnancies that attempted to deliver vaginally but later changed to C-section because of some anomaly found during labour- and urgent C-sections -in which no labour was tried.

From the 4,413 cases, I dropped from the analysis 115 of births that had missing information in at least one of the variables used. In addition, I keep only one observation per pregnancy and drop 396 observations from plural births. Then I restrict the sample to surgeons who have performed at least 6 C-sections a year, keeping 55 surgeons who performed 3,623 (92%) surgeries.[15]

[13] Certificato di assistenza al parto (CEDAP).

[14] The data in hand encompasses only births, hence I am blind to any other activities gynecologists may perform when not doing C-sections. Other surgeries include removal of the uterus, tubes and ovaries in the case of tumors, removal of ovarian cysts, removal of uterine fibroids, removal of "pathologic" tissue in endometriosis, treatment of ectopic pregnancies (ie, where the fetus develops out from the uterus), and more. Hockenberry and Helmchen (2014) utilize two measures of temporal break, time since the last CABG performed and time since any surgical procedure, and find that the last one affects patient outcomes substantially more than the procedure-specific measure. If that were the case also for surgeons performing C-sections, the estimates reported below would be biased towards zero and constitute a lower bound of the true effect.

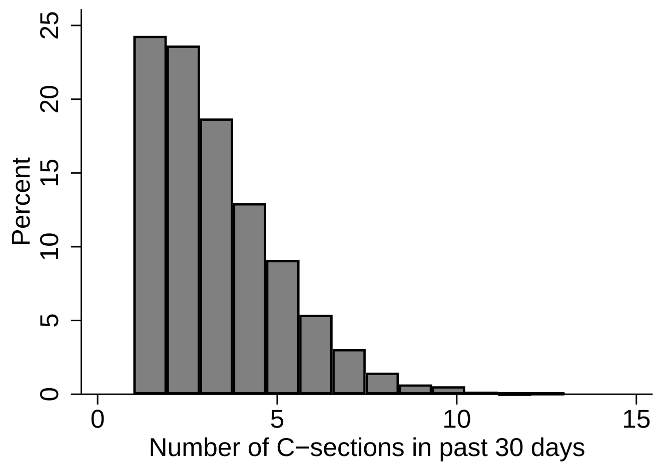
[15] Initially there were 85 surgeons in the sample, but many of those performed few and sporadic surgeries and were not registered again in my data.

Finally I drop the cases in which the number of days between a surgeon's last C-section and the current one is above 30 days (751 cases) because there may be other factors affecting their recent experience (e.g. surgeons who are just starting, sickness or paternal leave, etc). The study sample has 3,013 births performed by 55 surgeons in the 4 years.

Table 2.1 summarizes the variables used in the analysis for this sample. Mean admission to NICU was 21.3%[16], mean low Apgar score was 12.1%, and mean postoperative length of stay was 92.2 hours. For surgeons performing at least 20 C-sections a year, mean admission to NICU was 24.4%, mean low Apgar score was 13.8%, and mean postoperative length of stay was 92.8 hours, consistent with the hypothesis that high-volume surgeons take high-risk cases. Emergently admitted patients (in-labor C-sections) had higher probability of having a low Apgar score, but less likely to be admitted to NICU than elective patients; on the other hand, urgent cases have a higher probability of both bad outcomes.

The mean number of procedures performed in the previous 30 days was 3 for all surgeons, and 3.6 for high-volume surgeons. Surgeons performing emergent or urgent C-sections have a higher mean recent experience than those performing elective procedures. Figure 2.1 shows the distribution of the measure of recent experience for the whole sample.

Figure 2.1 – Distribution of cesarean sections by surgeon's recent experience.



The average age of patients is 34.5, and about 41% of them are first-time mothers -although this number is only 31% for elective procedures, and higher for the others. Similarly, about 20% of all births are born with less than 37 weeks of gestation, but they constitute 32% of urgent C-sections, and only 12.6% of emergent C-sections.

[16]This includes both intensive and sub-intensive units.

Table 2.1 – Summary Statistics

	Procedures performed by surgeons with at least 6 CS a year				Procedures performed by surgeons with at least 20 CS a year			
	All	Elective	In-labour	Urgent	All	Elective	In-labour	Urgent
Outcomes								
Neonatal Intensive Care Unit	21.3	18.1	16.9	31.2	24.4	22.1	18.2	33.7
Apgar Score below 9	12.1	8.1	12.9	17.9	13.8	8.5	13.1	19.5
Hours in hospital after surgery	92.2	89.7	93.5	95.0	92.8	90.9	93.2	94.3
Surgeon characteristics								
Num. C-sections in last 30 days	3.0	2.8	3.3	3.2	3.6	3.4	3.7	3.7
Days since last C-section	8.9	10.0	7.7	8.5	7.9	9.0	7.1	7.8
Mother characteristics								
Patient age (years)	34.5	35.2	33.6	34.2	34.3	35.5	33.6	34.0
Patient with university degree	31.3	32.9	31.2	28.9	32.1	38.2	30.6	28.0
First-time mother	41.4	31.2	54.2	44.1	43.5	31.4	54.3	42.9
Pregnancy characteristics								
Average weeks of gestation	37.7	37.7	38.6	36.9	37.6	37.4	38.5	36.7
Preterm birth (<37 weeks)	20.9	19.7	12.6	31.9	23.0	22.9	13.2	34.1
ER visit during pregnancy	19.6	22.3	15.4	20.0	19.5	23.5	15.4	20.3
Male newborn	51.5	50.2	55.7	49.2	52.7	53.7	54.5	49.8
Weight at birth	2,985.4	2,988.6	3,168.7	2,782.0	2,950.6	2,943.1	3,140.9	2,743.0
Low-weight-at-birth (<2.5 kilos)	21.2	20.4	13.3	31.0	22.7	22.9	14.3	32.2
Observations	3,013	1,312	884	817	1,614	503	589	522
Proportion	100.0	43.5	29.3	27.1	53.6	31.2	36.5	32.3

2.4.4 Outcomes

The most common outcome (almost exclusively) used in the health economics literature analyzing learning-by-doing and forgetting by hospitals and physicians is the death of the patient -both during and after surgery. As mentioned before, one important drawback of the database used here is the small sample size. Both maternal and fetal deaths are rare events, more so in developed countries, hence there are very few observations experiencing either one of these outcomes (e.g. there are only 12 stillbirths in the study sample). This impedes their use as outcomes for this study. However, one may also argue that mortality alone, being an extreme outcome, is an inadequate measure for capturing the full spectrum of the effects of learning-by-doing on patient health and hospital costs (e.g. morbidity or ordered procedures may also be important outcomes).

The data in hand contains other potential outcomes for patients' health beyond death that can be affected by surgeons skills. As proxies for newborns' health, this study uses the probability of needing to be transferred to a neonatal intensive care unit (NICU) and probability of having a low APGAR score. The first one measures whether the newborn had to be transferred to an NICU. The idea here is that with two equally healthy pregnancies, if one ends up going to NICU and the other doesn't, then something was done wrong -or not that well- in the first case. Furthermore, NICU admissions are among the most expensive treatments in regular hospitals, with one day cost being above $3,000. The second outcome is based on a total score of 1 to 10, the higher the score, the better the baby is doing after birth. This test is done to determine whether a newborn needs help breathing or is having heart trouble. Any score lower than 7 is a sign that the baby needs medical attention. In this study, there are only 72 newborns with score below 7. For this reason a new measure was constructed setting the bar higher and all births with a score lower than 9 will be considered of lower health. This doesn't necessarily mean a bad score that doctors should act on, but it can be argue that a newborn with an APGAR score below 9 is in worse health condition than a newborn with a score of 9 or 10. With regard to mothers' health, the only outcome available that can be affected by surgeons skills is the postoperative length of stay in hospital (PLOS), measured in number of days.

2.5 Results

2.5.1 The effect of recent practice on patient health

Table 2.2 presents the average marginal effects of recent experience in a linear probability model.[17][18] Panel (A) shows coefficients for experience when pulling together all C-sections, while Panel (B) shows coefficients by type of C-section. The first three columns are estimated using all physicians, while the last three use only those performing at least 20 C-sections a year (high volume surgeons).

Table 2.2 – Average marginal effect of experience on outcomes

	Surgeons performing at least 6 C-sections a year			Surgeons performing at least 20 C-sections a year		
	NICU§	Apgar<9	PLOS‡	NICU§	Apgar<9	PLOS‡
Panel A:						
All	-0.005	-0.005	0.002	-0.009**	-0.009**	0.005
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Panel B:						
Elective	0.003	-0.002	0.007	-0.004	-0.003	0.008
	(0.005)	(0.004)	(0.005)	(0.007)	(0.006)	(0.007)
In-labor	-0.012**	-0.011**	-0.002	-0.012*	-0.014**	0.002
	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.007)
Urgent	-0.008	-0.003	-0.000	-0.010	-0.009	0.005
	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)
Obs.	3,013	3,013	2,884	1,614	1,614	1,532
Mean Dep.	0.213	0.121	4.462	0.220	0.110	4.510

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$. Reported coefficients are average marginal effects. Panel (A) presents coefficients for experience when using the full sample, and Panel (B) are by type of C-section. §NICU: Neonatal Intensive Care Unit; ‡PLOS: Postoperative length of stay (in log-hours). All models include controls for year, month and day of the week fixed effects, surgeon fixed effects and controls for emergency visits during pregnancy, low weight-at-birth, a quadratic term for weeks-of-gestation, newborn's gender, a quadratic term for mother's age, mother's education and whether she is first-time mother.

If one were to focus only on the full sample of C-sections and physicians, reported coefficients suggest that there is very small and insignificant effects

[17] Table B.2 compares results using a probit or a linear probability model and coefficients are virtually the same up to the second decimal.

[18] In Appendix B.1 I provide a more detailed discussion on model selection, where models of different polynomial degrees and categorical definitions of recent-experience are tested. A linear specification seems to be the one that better fits the data.

of recent experience on outcomes, or in other words, very little surgeon's skills depreciation. However, for patients whose whose C-section was done by a high-volume surgeon (Panel A, last three columns), an additional C-section in the previous 30 days was associated with a decrease in the likelihood of both NICU admission and low Apgar-Score by nearly 1 percentage point (or 3.7% and 6.5% relative to the sample means). Similar to results on the full-sample of surgeons, an additional C-section in the past 30 days has no statistically nor economically significant effect on the postoperative length of stay.

The estimated impact of recent experience on outcomes was even larger when looking at emergent cases. For physicians performing an in-labour C-section, having performed one additional C-section in the previous 30 days was associated with a decrease in neonatal intensive care unit admission by nearly 1.2 percentage points (or 5.5% relative to the sample mean) and a decrease in low Apgar Score of about 1.1 percentage points (or 10% relative to the sample mean). The effect was smaller in magnitude and not statistically significant for the case or urgent or elective C-sections. Once again, the effects on postoperative length of stay were very precisely estimated close to zero for all three types of surgeries. Surprisingly, unlike the before when looking at all C-sections together, results for high-volume surgeons by type of C-sections are quite similar to those for the whole sample of surgeons.

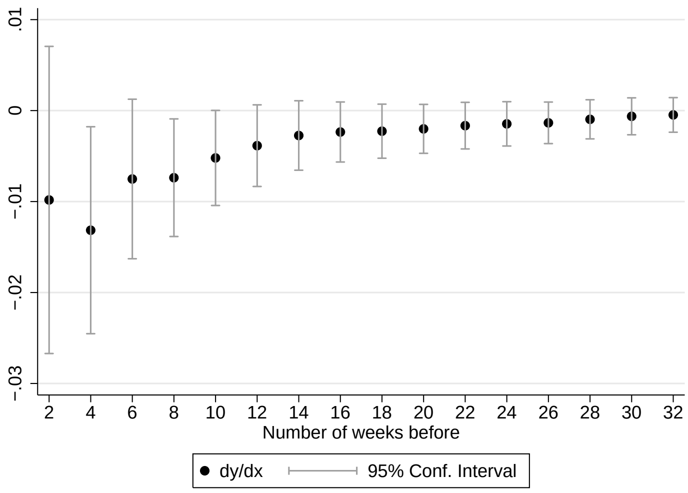
2.5.2 Robustness checks

A priori, there is no clear criteria to choose a specific time period for my measure of recent experience. If one were to choose a very long period, it could happen that the effect of the further away surgeries have little impact on today's one. On the other hand, choose a period too short and maybe there is not enough variation in the amount of experience. In this paper I decided to measure recent experience within the last 4 weeks (30 days). To test how sensitive results are to this decision, I run a set of regression for different time spams (from 2 weeks up to 32 weeks) for the three outcomes. Figure 2.2 shows the results for the case of (emergent) in-labor C-sections using the full sample of surgeons. For both NICU admission and low Apgar Score, the effect of the number of previous C-sections gets monotonically smaller the longer is the measurement period (except when using a time period of two, although here the standard errors are considerably higher). This provides further evidence for the human capital depreciation hypothesis, where procedures performed more than one year back have little effect on surgeon's ability today -controlling for her average ability. Similar to previous findings, there is a consistent not significant effect of recent experience on the mother's postoperative length of stay across time windows.

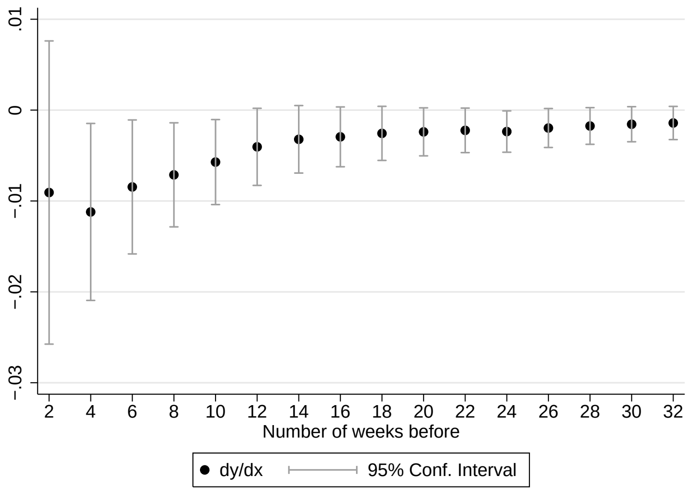
To test whether selection bias affects these results, I regress each pre-treatment characteristic on the treatment -the number of C-sections performed in the last 30 days-. If observed characteristics were associated with recent experience,

Figure 2.2 – Average marginal effect of recent experience measured in different time windows.

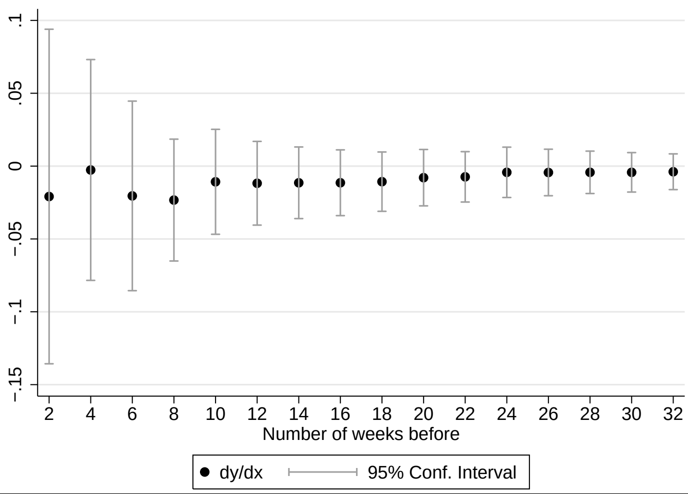
(a) Neonatal Intensive Care Unit



(b) Apgar Score < 9



(c) Postoperative Length of Stay



it would be a sign of patient selection. The results for these estimations are reported in Table B.3. All coefficients are small in absolute and relative size, and all but one are not statistically different from zero. The only case that it is significant at the 90% level is for having had an emergency visit during pregnancy, but the point estimate implies only a 4.5% change.

2.6 Discussion

There is a well established positive association in healthcare between providers' volume and health outcomes, yet our current understanding on the drivers behind this correlation are limited. The two leading explanatory mechanisms are 'learning-by-doing' and 'selective referral'. In this paper I use a unique feature of the Italian health care system -patients are not allowed to choose physician- to investigate whether there is evidence of 'learning-by-doing' in cesarean section surgeons. More specifically, I test whether surgeons who have performed more procedures in the recent past observe an improvement in performance. The contribution is threefold: First it my empirical approach rests on an institutional context that allows me to estimate parameters that are free of the self-selection bias that may arise when patients can sort across physicians. Second, I provide some evidence that learning-by-doing effects are heterogeneous across procedure types depending on their emergent nature. Finally, I investigate this for C-sections, a procedure that is nowadays very relevant but has been ignored so far in the literature of learning-by-doing.

Using information on birth certificates for one large hospital in Italy during between 2011 and 2014, I find that, for emergent cases, performing one additional procedure reduces the likelihood of neonatal intensive care unit admission by nearly 1.2 percentage points (5.5%) and of being born with a low Apgar Score by about 1.1 percentage points (10%), all else equal. This effect is not present for the case of elective C-sections.

In terms of policy recommendation, the results of this study would add to the benefits of centralizing cesarean section services in a small number of bigger maternity wards. The evidence here suggests that physicians who perform more C-sections in the recent past are better when executing the next surgery, specially in the cases where it is an emergency C-section.

Appendix A

A.1 The working sample and scheduled patients

The working sample used in the main paper is restricted to only those unscheduled patients who attempt to have a vaginal delivery after going through labor and leaves out scheduled patients. Scheduled patients can be further divided in two groups: (i) elective C-sections, and (ii) pharmacologically-induced patients. This appendix shows evidence of how the latter group's transition through the maternity ward resembles more that of elective C-section rather than the one of unscheduled patients, and hence should not be included in the working sample.

One important caveat of the data is that one cannot disentangle scheduled from unscheduled patients among those who were pharmacologically induced. However, anecdotal evidence from the ward's staff suggest that most of them are scheduled (e.g. overdue pregnancy). Furthermore, a descriptive analysis of the data seems to corroborate that. Figures A.1 and A.2 present the distribution of patients across hours and days as performed in section 1.3.1 of the main paper except that now scheduled patients are further divided between elective C-sections and induced. Starting from Figure A.1, it shows that there is a pick in admissions for both elective C-sections and induced patients during the afternoon shift, and then again a pick in time of birth (although the pick is later in the day for induced patients relative to the elective C-sections). Nevertheless, the picks are less pronounced for induced patients, suggesting that some of them may be arriving at random hours of the day like unscheduled patients do.

Even though the distribution by hours of induced patients seem to follow that of elective C-sections, their distribution by day of the week instead is closer to that of unscheduled patients. Even though admissions are slightly lower during weekends, births are evenly distributed across all days of the week. This is probably due to the fact that, as long as everything goes well, these patients are taken care of by midwives (not physicians).

The evidence provided in this appendix supports the idea of excluding both elective C-sections and pharmacologically induced patients from the working sample, but to include the latter group in the treatment variable given that they are primordially seen by midwives.

Figure A.1 – Distribution of admissions and births by hour.

(a) By Time of Admission **(b)** By Time of Birth

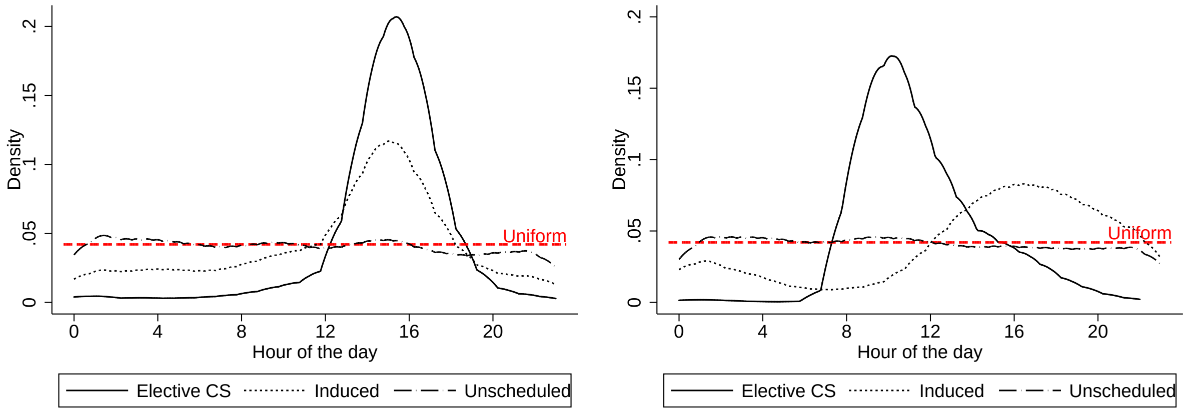
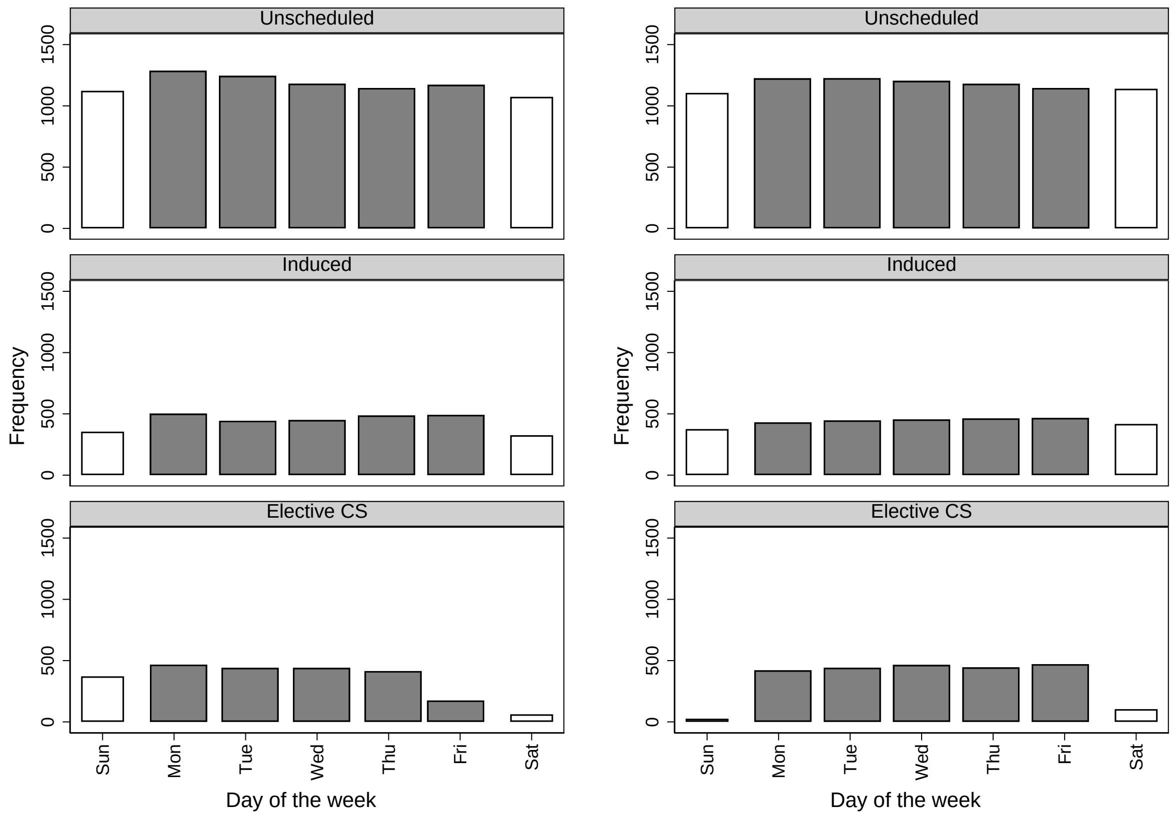


Figure A.2 – Frequency of admissions and births by day.

(a) By Day of Admission **(b)** By Day of Birth



A.2 A measure of workload without adjusting for supply side factors

The measure of workload used in the main paper is the ratio of patients-to-midwives, hence it takes into account both demand and supply side effects. Specifying the covariate of interest as a ratio may put some constraints on the estimated coefficient. This appendix repeats the main estimations but using instead the number of unscheduled patients waiting to give birth (without adjusting for the number of midwives).

Figures A.3 shows a histogram of the number of unscheduled patients observed by each patient at admission. The mode is 3, and the mean is slightly above at 3.34. As in the main paper, I divide this variable in quintiles to test for non-linearities in its effect on the probability of C-section. Table A.1 describes the number of observations and limits for each quintile.

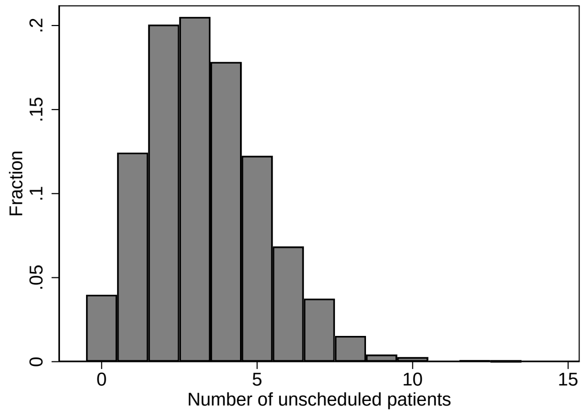


Figure A.3 – Histogram unscheduled patients

Quintile	N	Min	Max
1	953	0	2
2	536	3	3
3	466	4	4
4	320	5	5
5	338	6	13

Table A.1 – Descriptive statistics of quintiles

Finally Table A.2 presents the results from running the preferred model using the number of unscheduled patients as regressor. Similar to the findings in the main paper, there seems to be a non-linear relationship between workload and the probability of C-section. This effect starts to rise already in the second quintile and slowly declines in the fourth and fifth quintiles. This provides more assurance to the results using the ratio of patients-to-midwives.

A.3 Robustness to alternative models

In the main paper two functional forms are tested for the effect of workload on the probability of C-section: a linear specification, and a non-linear one using a categorical variable constructed from the 20th and 80th percentiles. This appendix elaborates further on the model selection and tests other specifications. Columns (1) to (4) in Table A.3 present the coefficients for different

Table A.2 – Average marginal effect on probability of C-section

	Linear	Non-linear
Number of unscheduled patients	0.0044	
	(0.0036)	
2nd Quintile		0.0309*
		(0.0177)
3rd Quintile		0.0503**
		(0.0200)
4th Quintile		0.0384*
		(0.0223)
5th Quintile		0.0266
		(0.0212)
Observations	2,613	2,613
Mean dep.	0.119	0.119

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

polynomial degrees of the ratio of patients-to-midwives, with the Akaike Information Criteria (AIC) reported at the bottom. It seems that, within these polynomial functional forms, the data at hand is better represented by either a squared or cubic polynomial, given their statistical significance and their low AIC.

Column five presents results using a categorical variable with the quintiles of the distribution of the ratio of patients-to-midwives (where the reference group is the first quintile). This specification gives the model more flexibility to fit the data, at the cost of estimating more coefficients. Results suggest that there is a sudden rise in the probability of C-section for patients who see a ratio of patients-to-midwives in the second quintile, which then falls slowly until the fifth quintile where it is no longer statistically distinguishable from the reference group. This decay in the probability of C-section for higher workloads may be associated with capacity constraints on the operative theater (beds, number of gynecologists, etc.).

Given the previous , I created a variable with three categories where the 3 middle quintiles of the ratio of patients-to-midwives have been coded together in one group (<20th percentile, 20-80th percentile, >80th percentile). This specification has the advantage of capturing the higher level of C-sections that occurs in the middle of the workload distribution, while diminishing the number of coefficients to be estimated and augmenting precision. Results are presented in the sixth column.

Table A.3 – Alternative model specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Ratio	0.0004	0.0524**	0.1195**	0.1627		
	(0.0085)	(0.0240)	(0.0535)	(0.1283)		
Ratio Square		-0.0104**	-0.0369**	-0.0633		
		(0.0044)	(0.0182)	(0.0719)		
Ratio Cubic			0.0030*	0.0091		
			(0.0018)	(0.0157)		
Ratio Quadratic				-0.0005		
				(0.0011)		
2nd Quintile					0.0554***	
					(0.0199)	
3rd Quintile					0.0329*	
					(0.0196)	
4th Quintile					0.0299	
					(0.0211)	
5th Quintile					0.0167	
					(0.0216)	
20-80th Percentile						0.0408***
						(0.0156)
>80th Percentile						0.0197
						(0.0213)
Observations	2,613	2,613	2,613	2,613	2,613	2,613
AIC	1544.59	1542.58	1543.18	1545.06	1541.89	1539.65

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

A.4 Other Graphs and Tables

Figure A.4 – Density distribution of Ratio

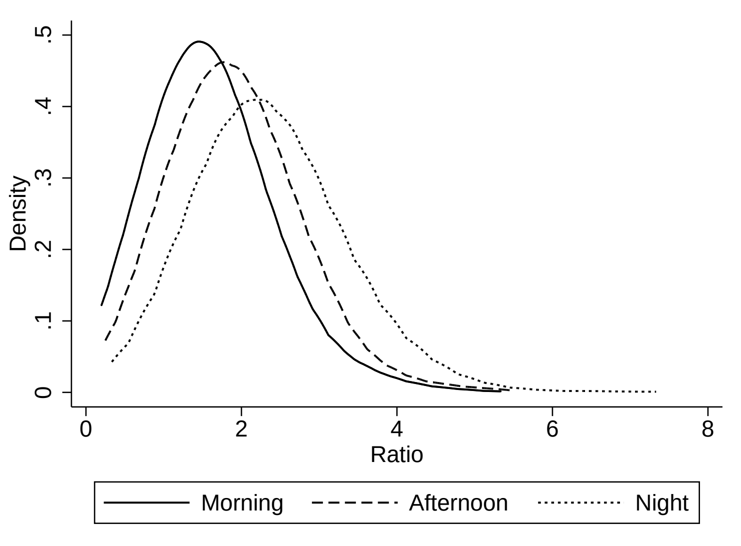


Figure A.5 – Ratio by time of admission

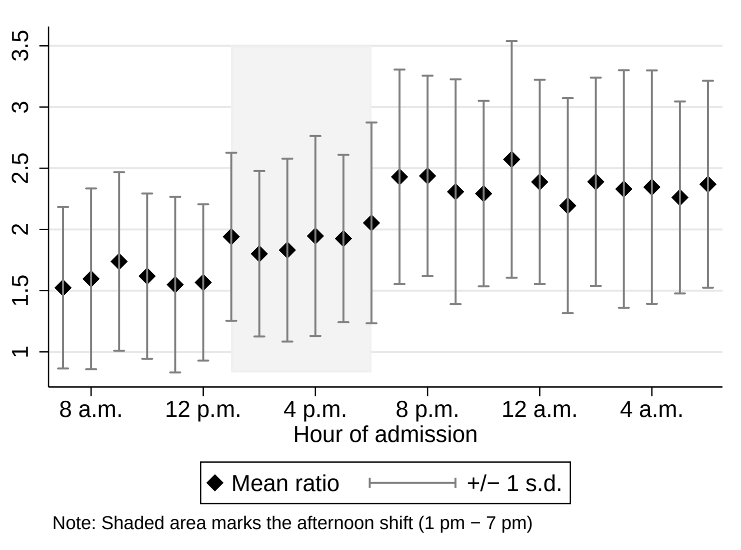


Table A.4 – Regression of pre-treatment characteristics on Ratio of patients-to-midwives

Dependent variable	Coef. of Ratio
Mother's characteristics	
With university degree	-0.0141
	(0.0129)
Above 36 yo	-0.0033
	(0.0117)
Pregnancy's characteristics	
Preterm (before 37th week)	-0.0080
	(0.0057)
At least 1 ER visit	-0.0005
	(0.0076)
Newborn's characteristics	
Male	-0.0045
	(0.0132)
Low weight at birth	-0.0054
	(0.0055)
Observations	2,613

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table A.5 – Pre-treatment variables balanced across treatments and control

	Level of Ratio				
	<20th Percentile	20-80th Percentile	>80th Percentile	(1) vs. (2)	(1) vs. (3)
Mother's characteristics					
% of mother's with university degree	0.386	0.358	0.340	0.271	0.112
	(0.022)	(0.013)	(0.018)		
% older than 36 yo	0.302	0.288	0.266	0.567	0.182
	(0.021)	(0.012)	(0.017)		
Pregnancy's characteristics					
% of births before 37 weeks of gestation	0.057	0.058	0.040	0.942	0.191
	(0.011)	(0.006)	(0.008)		
% of pregnancies with at least 1 ER visit	0.110	0.123	0.100	0.431	0.597
	(0.014)	(0.009)	(0.012)		
Newborn's characteristics					
% of male newborns	0.508	0.517	0.497	0.755	0.704
	(0.023)	(0.013)	(0.019)		
% of low-weight newborns (<2,500 grams)	0.055	0.051	0.042	0.746	0.305
	(0.010)	(0.006)	(0.008)		

Standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table A.6 – Probability of C-section using a Linear Probability and Probit Model

	LPM	Probit
Panel (A):		
Ratio patients	0.0007	0.0012
to midwives	(0.0083)	(0.0081)
Panel (B):		
20-80th Percentile	0.0410***	0.0437***
	(0.0154)	(0.0166)
>80th Percentile	0.0187	0.0222
	(0.0210)	(0.0221)
Observations	2,613	2,613
Mean dep.	0.119	0.119

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

Table A.7 – LPM of C-section by day and staff shift

	All	Weekday	Weekend	Morning	Afternoon	Night
20-80th Percentile	0.0410***	0.0377**	0.0608*	0.0467*	0.0341	0.0242
	(0.0154)	(0.0176)	(0.0325)	(0.0270)	(0.0300)	(0.0245)
>80th Percentile	0.0187	0.0195	0.0251	0.0054	0.0477	0.0014
	(0.0210)	(0.0245)	(0.0429)	(0.0574)	(0.0523)	(0.0282)
Observations	2,613	1,883	730	636	641	1,336
Mean dep.	0.119	0.114	0.132	0.104	0.137	0.118

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table A.8 – Effect of staffing for different windows of time since admission.

	1 hour	2 hours	3 hours	4 hours	5 hours	6 hours
20-80th Percentile	0.0408***	0.0490***	0.0469***	0.0399**	0.0308*	0.0241
	(0.0156)	(0.0156)	(0.0163)	(0.0162)	(0.0164)	(0.0164)
>80th Percentile	0.0197	0.0213	0.0320	0.0136	0.0100	0.0077
	(0.0213)	(0.0209)	(0.0222)	(0.0220)	(0.0217)	(0.0216)
Observations	2,613	2,613	2,613	2,613	2,613	2,613

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table A.9 – Effect of staffing at admission and 24hs after

	Workload at admission		Workload 24hs after admission	
Panel (A)				
Ratio patients	0.0019	0.0004	0.0020	-0.0010
to midwives	(0.0074)	(0.0085)	(0.0073)	(0.0081)
Panel (B)				
20-80th Percentile	0.0432***	0.0408***	-0.0019	-0.0065
	(0.0149)	(0.0156)	(0.0160)	(0.0161)
>80th Percentile	0.0193	0.0197	0.0011	-0.0091
	(0.0196)	(0.0213)	(0.0202)	(0.0213)
Time FE	✓	✓	✓	✓
Controls		✓		✓
Other patients		✓		✓
Shift*DOW FE		✓		✓

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Appendix B

B.1 Robustness to alternative models

Because there might be non-linearities in the effect of recent-experience on health outcomes, this appendix repeats the main estimation for different specifications of the covariate of interest. The analysis is performed only on the sample of high-volume surgeons. Columns (1) to (3) in Table B.1 present the coefficients including different polynomial degrees of recent-experience, and column (4) divides the distribution in three using percentiles (below 33th percentile, between 33-66th percentiles, and above 66th percentile). At the bottom of the table I report the Akaike Information Criteria (AIC). The same exercise is repeated for the three outcomes analyzed in the paper.

For all cases, a linear specification seems to be the one that better fits the data, although when looking at the probability of NICU admission there is some evidence that the effect is significant only for very high values of recent-experience.

Table B.1 – Alternative model specifications

	NICU§				Apgar<9				PLOS‡			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Experience	-0.009**	-0.016	-0.059**		-0.009**	-0.008	0.015		0.005	0.009	0.025	
	(0.004)	(0.012)	(0.028)		(0.004)	(0.012)	(0.027)		(0.004)	(0.013)	(0.029)	
Exp. Square		0.001	0.011*			-0.000	-0.005			-0.000	-0.004	
		(0.001)	(0.006)			(0.001)	(0.005)			(0.001)	(0.006)	
Exp. Cubic			-0.001*				0.000				0.000	
			(0.000)				(0.000)				(0.000)	
33-66th percentile				-0.015				0.006				0.027
				(0.024)				(0.024)				(0.024)
>66th percentile				-0.045*				-0.024				0.025
				(0.026)				(0.027)				(0.028)
Observations	1,614	1,614	1,614	1,614	1,614	1,614	1,614	1,614	1,532	1,532	1,532	1,532
AIC	804.2	805.8	805.2	806.3	681.1	683.1	684.3	684.5	864.9	866.8	868.5	866.7

§NICU: Neonatal Intensive Care Unit. Robust standard errors in parentheses. ‡PLOS: Postoperative length of stay (in log-hours). $***p < 0.01, **p < 0.05, *p < 0.1$. Reported coefficients are average marginal effects. All models include controls for year, month and day of the week fixed effects, surgeon fixed effects and controls for emergency visits during pregnancy, low weight-at-birth, a quadratic term for weeks-of-gestation, newborn's gender, a quadratic term for mother's age, mother's education and whether she is first-time mother. The sample constitutes only high-volume surgeons.

B.2 Other graphs and tables

Table B.2 – Linear Probability vs. Probit Model

	Probit		LPM	
	NICU§	Apgar<9	NICU§	Apgar<9
Experience	-0.006*	-0.006*	-0.005	-0.005
	(0.003)	(0.003)	(0.004)	(0.003)
Obs.	3,013	2,947	3,013	3,013
Mean Dep.	0.213	0.124	0.213	0.121

§NICU: Neonatal Intensive Care Unit. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$. Reported coefficients are average marginal effects. All models include controls for year, month and day of the week fixed effects, surgeon fixed effects and controls for emergency visits during pregnancy, low weight-at-birth, a quadratic term for weeks-of-gestation, newborn's gender, a quadratic term for mother's age, mother's education and whether she is first-time mother.

Table B.3 – Regression of pre-treatment characteristics on recent experience

	University degree	Mother's age	First-time mother	Male newborn	Low weight at birth	Preterm birth	Emergency visit
Experience	0.001	0.009	0.000	0.002	0.003	-0.002	0.008
	(0.006)	(0.066)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Obs.	3,013	3,013	3,013	3,013	3,013	3,013	3,013
Mean dep.	0.313	34.46	0.414	0.515	0.212	0.209	0.196

Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$. Reported coefficients are average marginal effects. All models include controls for year, month and day of the week fixed effects and surgeon fixed effects.

Bibliography

Diane Alexander et al. Does physician compensation impact procedure choice and patient health? Technical report, 2013.

Unzila A Ali and Errol R Norwitz. Vacuum-assisted vaginal delivery. *Reviews in Obstetrics and Gynecology*, 2(1):5, 2009.

Sara Allin, Michael Baker, Maripier Isabelle, and Mark Stabile. Physician incentives and the rise in c-sections: Evidence from canada. Technical report, National Bureau of Economic Research, 2015.

American College of Obstetricians and Gynecologists. Faqs: Cesarean birth. Technical report, 2010. URL `http://www.acog.org/~/media/For%20Patients/faq006.pdf?dmc=1&ts=20120731T1617495597`.

Ramji Balakrishnan and Naomi S Soderstrom. The cost of system congestion: Evidence from the healthcare sector. *Journal of Management Accounting Research*, 12(1):97–114, 2000.

Donald M Berwick and Andrew D Hackbarth. Eliminating waste in us health care. *Jama*, 307(14):1513–1516, 2012.

John D Birkmeyer, Andrea E Siewers, Emily VA Finlayson, Therese A Stukel, F Lee Lucas, Ida Batista, H Gilbert Welch, and David E Wennberg. Hospital volume and surgical mortality in the united states. *New England Journal of Medicine*, 346(15):1128–1137, 2002.

Jan Blustein and Jianmeng Liu. Time to consider the risks of caesarean delivery for long term child health. *BMJ: British Medical Journal*, 350, 2015.

Juan M Contreras, Beomsoo Kim, and Ignez M Tristao. Does doctors' experience matter in lasik surgeries? *Health economics*, 20(6):699–722, 2011.

Andrew Cook, Martin Gaynor, Melvin Stephens Jr, and Lowell Taylor. The effect of a hospital nurse staffing mandate on patient health outcomes: Evidence from california's minimum staffing regulation. *Journal of Health Economics*, 31(2):340–348, 2012.

Janet Currie and W Bentley MacLeod. First do no harm? tort reform and birth outcomes. *Quarterly Journal of Economics*, 123(2), 2008.

SC Curtin, KD Gregory, LM Korst, and SF Uddin. Maternal morbidity for vaginal and cesarean deliveries, according to previous cesarean history: New data from the birth certificate, 2013. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 64(4):1, 2015.

Pedro de Araujo, Maroula Khraiche, and Andrea Tukan. Does overcrowding and health insurance type impact patient outcomes in emergency departments? *Health Economics Review*, 3(1):1–7, 2013.

Catherine Deneux-Tharaux, Elodie Carmona, Marie-Hélene Bouvier-Colle, and Gérard Bréart. Postpartum maternal mortality and cesarean delivery. *Obstetrics & Gynecology*, 108(3, Part 1):541–548, 2006.

David Dranove and Yasutora Watanabe. Influence and deterrence: How obstetricians respond to litigation against themselves and their colleagues. *American Law and Economics Review*, 2009.

William N Evans and Beomsoo Kim. Patient outcomes when hospitals experience a surge in admissions. *Journal of Health Economics*, 25(2):365–388, 2006.

Maura Francese, Massimiliano Piacenza, Marzia Romanelli, and Gilberto Turati. Understanding inappropriateness in health spending: The role of regional policies and institutions in caesarean deliveries. *Regional Science and Urban Economics*, 49:262–277, 2014.

Martin Gaynor, Harald Seider, and William B Vogt. The volume-outcome effect, scale economies, and learning-by-doing. *American Economic Review*, pages 243–247, 2005.

Kimberly D Gregory, Sherri Jackson, Lisa Korst, and Moshe Fridman. Cesarean versus vaginal delivery: whose risks? whose benefits? *American Journal of Perinatology*, 29(01):07–18, 2012.

Jon Gruber, John Kim, and Dina Mayzlin. Physician fees and procedure intensity: the case of cesarean delivery. *Journal of Health Economics*, 18(4): 473–490, 1999.

Jonathan Gruber and Maria Owings. Physician financial incentives and cesarean section delivery. *The Rand Journal of Economics*, pages 99–123, 1996.

Jostein Grytten, Irene Skau, and Rune Sørensen. Do expert patients get better treatment than others? agency discrimination and statistical discrimination in obstetrics. *Journal of Health Economics*, 30(1):163–180, 2011.

Ethan A Halm, Clara Lee, and Mark R Chassin. Is volume related to outcome in health care? a systematic review and methodologic critique of the literature. *Annals of internal medicine*, 137(6):511–520, 2002.

Marie Hatem, Jane Sandall, Declan Devane, Hora Soltani, and Simon Gates. Midwife-led versus other models of care for childbearing women. *The Cochrane Library*, 2008.

V. Ho. Learning by doing. In Anthony J. Culyer, editor, *Encyclopedia of Health Economics*, pages 141 – 145. Elsevier, San Diego, 2014. ISBN 978-0-12-375679-4. doi: http://dx.doi.org/10.1016/B978-0-12-375678-7.01110-X. URL http://www.sciencedirect.com/science/article/pii/B978012375678701110X.

Vivian Ho. Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. *Journal of health economics*, 21(5):873–885, 2002.

Jason M Hockenberry and Lorens A Helmchen. The nature of surgeon human capital depreciation. *Journal of health economics*, 37:70–80, 2014.

Robert S Huckman and Gary P Pisano. The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science*, 52(4): 473–488, 2006.

Marco D Huesch. Learning by doing, scale effects, or neither? cardiac surgeons after residency. *Health services research*, 44(6):1960–1982, 2009.

Marco D Huesch and Mariko Sakakibara. Forgetting the learning curve for a moment: how much performance is unrelated to own experience? *Health Economics*, 18(7):855–862, 2009.

Stanley Ip, Mei Chung, Gowri Raman, Priscilla Chew, Nombulelo Magula, Deirdre DeVine, Thomas Trikalinos, and Joseph Lau. Breastfeeding and maternal and infant health outcomes in developed countries. *Evidence report/technology assessment*, pages 1–186, 2007.

Erin M. Johnson and M. Marit Rehavi. Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, 8(1):115–41, 2016. doi: 10.1257/pol.20140160. URL http://www.aeaweb.org/articles.php?doi=10.1257/pol.20140160.

Emmett B Keeler and Mollyann Brodie. Economic incentives in the choice between vaginal delivery and cesarean section. *The Milbank Quarterly*, pages 365–404, 1993.

Joy E Lawn, Judith Mwansa-Kambafwile, Bernardo L Horta, Fernando C Barros, and Simon Cousens. Kangaroo mother care to prevent neonatal deaths due to preterm birth complications. *International Journal of Epidemiology*, 39(suppl 1):i144–i154, 2010.

Ann G Lawthers, Nan M Laird, Stuart Lipsitz, Liesi Hebert, Troyen A Brennan, and A Russell Localio. Physicians' perceptions of the risk of being sued. *Journal of Health Politics, Policy and Law*, 17(3):463–482, 1992.

Mélanie Lefèvre. Physician induced demand for c-sections: does the convenience incentive matter? 2014.

Millard F Long and Paul J Feldstein. Economics of hospital systems: peak loads and regional coordination. *The American Economic Review*, 57(2): 119–129, 1967.

DN Lucas, SM Yentis, SM Kinsella, A Holdcroft, AE May, M Wee, and PN Robinson. Urgency of caesarean section: a new classification. *Journal of the Royal Society of Medicine*, 93(7):346–350, 2000.

Harold S Luft, John P Bunker, and Alain C Enthoven. Should operations be regionalized? the empirical relation between surgical volume and mortality. *New England Journal of Medicine*, 301(25):1364–1369, 1979.

Harold S Luft, Sandra S Hunt, and Susan C Maerki. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health services research*, 22(2):157, 1987.

Elizabeth R Moore, Gene C Anderson, Nils Bergman, et al. Early skin-to-skin contact for mothers and their healthy newborn infants. *Cochrane Database Syst Rev*, 3, 2007.

Deirdre J Murphy, Maureen Macleod, Rachna Bahl, and Bryony Strachan. A cohort study of maternal and neonatal morbidity in relation to use of sequential instruments at operative vaginal delivery. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 156(1):41–45, 2011.

Karen Norberg and Juan Pantano. Cesarean sections and subsequent fertility. *Journal of Population Economics*, 29(1):5–37, 2016.

World Health Organization. *The world health report 2000: health systems: improving performance.* World Health Organization, 2000.

Osservatorio di Epidemiologia dell'Agenzia Regionale di Sanit (ARS) della Toscana. Nascere in toscana. anni 2008-2011. Technical report, 2013. URL `http://www.ars.toscana.it/files/pubblicazioni/Volumi/2013/72_cap_2013.pdf`.

Anne Pfuntner, LM Wier, and C Stocks. Most frequent procedures performed in us hospitals, 2011: Statistical brief# 165. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*, 2013.

Subbu Ramanarayanan. Does practice make perfect: An empirical analysis of learning-by-doing in cardiac surgery. *Mimeo*, 2008.

Andrew Sfekas. Learning, forgetting, and hospital quality: an empirical analysis of cardiac procedures in maryland and arizona. *Health economics*, 18 (6):697–711, 2009.

Dena Towner, Mary Ames Castro, Elaine Eby-Wilkens, and William M Gilbert. Effect of mode of delivery in nulliparous women on neonatal intracranial injury. *New England Journal of Medicine*, 341(23):1709–1714, 1999.