Departament d´Economia Aplicada

Document de treball

Forgetting-by-not-doing:

The case of surgeons and cesarean sections.

Gabriel A. Facchini Palma

Universitat Autònoma de Barcelona

Facultat d´Economia i Empresa

20.10

Aquest document pertany al Departament d´Economia Aplicada

Data de publicació: **Octobre 2020**

Departament d´Economia Aplicada

Edifici B

Campus de Bellaterra

08193 Bellaterra

Telèfon: 00 34 935 811 680

E.mail: d.econ.aplicada@uab.cat

https://www.uab.cat/departament/economia-aplicada/

Forgetting-by-not-doing:

The case of surgeons and cesarean sections.

Gabriel Facchini*

October, 2020

Abstract

This paper provides new evidence on the link between patient outcome and physician experience. Using birth certificates data from a large hospital in Italy, I analyze 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 their physician, I lower concerns of potential reverse causality (selective referral). I find evidence indicating a strong learning-by-doing effect: for emergent cases, a one standard deviation increase in recent experience reduces the likelihood of neonatal intensive care unit admission by nearly 2.9 percentage points (12%) and of being born with a low Apgar Score by about 1.3 percentage points (9.5%), all else equal. This effect is not present for the case of elective C-sections.

JEL Classification: J24, I10, I18

Keywords: Learning-by-doing, human capital, experience, volume, cesarean section, productivity

*Department of Applied Economics, Universitat Autònoma de Barcelona, Spain. Email: gabriel.facchini@uab.cat. I would like to extend my gratitude to Andrea Ichino and Jérôme Adda for their invaluable support and advice throughout this research project. My gratitude goes to Joseph Doyle and Liertad González as well, their detailed comments have substantially improved this paper. I am extremely grateful to Carlo Dani, Simone Pratesi, Federico Mecacci, Franca Rusconi and Luigi Gagliardi for their clinical expertise, to Tommaso Lanis for performing the confidential merge, and to Francesca Superbi for her assistance in accessing the data. All errors remain my own.

1 Introduction

Since the seminal report by (Luft et al., 1979), there has been growing evidence of a positive association between volume and quality in the provision of health services for a wide variety of procedures, time periods, and locations.[1] Nevertheless, the debate about the causal direction of this relationship is far from settled (Halm et al., 2002; Ho, 2014).

Two principal hypothesis have been put forward to explain this association: (i) 'learning-by-doing' (or 'practice-makes-perfect') and (ii) 'selective referral' (Luft et al., 1987).[2] Under 'learning-by-doing', increased experience leads to improvement in skills which in turn results in better quality as measured by patient outcomes. 'Selective referral', instead, occurs when providers with higher quality attract a larger 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 raise quality. 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 and lower geographical coverage.

This paper aims at causally identifying whether learning-by-doing is present at the individual level in the healthcare sector, more specifically, for surgeons performing cesarean sections (C-sections). In particular, I look at whether a surgeon's recent procedure volume affects patient outcomes.[3] In order to establish a causal relationship, I benefit from the fact that, due to state regulation, most pregnant women in Italy do not choose the gynecologist that will help them give birth within the public system. This institutional feature creates a setup 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 surgery. Even though patients cannot choose a particular physician, the hospital may assign physicians with higher skills to patients with a higher health risk (selective allocation). To address this concern 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: operations performed by physicians with a higher recent experience result in better newborn health. More specifically, I find that a one standard deviation increase in surgeon's experience in the previous four weeks lowers

[1] See Halm et al. (2002) and Chowdhury et al. (2007) for a revision of the medical literature.

[2] For a comprehensive analysis of the different channels that can explain the association between volume and outcome in health care see Huesch and Sakakibara (2009).

[3] This type of volume-outcome relationship has been called learning-from-recent-experience by Huesch and Sakakibara (2009).

a newborn's probability of having a low Apgar score by 9.5%[4] and of being admitted to a neonatal intensive care unit (NICU) by 12%. These effects are 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.

One important assumption for these results to hold is the absence of any form of dynamic matching between physicians and patients. If, for example, hospitals aware of depreciating skills may assign healthier patients to physicians with a low recent activity, which would biased the estimates towards zero. In this case the results should be considered a lower bound of the true effect. To alleviate this concern, I show that pre-treatment pregnancy and mother characteristics are uncorrelated with physician's recent experience. In a robustness check I restrict the sample to physicians who have performed at least one C-section in the previous four weeks to make the sample of cases more comparable. Results are virtually the same.

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.[5] Furthermore, the discussion on volume-outcome effects become all the more relevant in view of the recent wave of closures of maternity services in various countries (e.g. US, Canada, UK, Japan, France, the Netherlands, and others).[6] To the best of my knowledge, this is the first paper to obtain causal estimates of learning-by-doing for the case of cesarean section surgeons.

Literature Review

There are hundreds of papers in the medical literature finding an association between higher hospital or surgeon procedure volume and better health outcomes (Halm et al., 2002; Birkmeyer et al., 2003; Chowdhury et al., 2007). However, these studies are mostly observational and tend to neglect the potentially endogenous nature of provider volume. Few studies have attempted to translate the association between volume and outcome into a causal relationship, and most rigorous econometric analysis have failed to identify learning-by-doing (Ho, 2014).

At the hospital-level, studies on learning-by-doing typically use lagged or cumulative

[4]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.

[5]In the US alone, in 2011 there were almost 1.3 million C-sections (Pfuntner et al., 2013).

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

volume as covariates of interest, and find no support for the learning-by-doing hypothesis (Gaynor et al., 2005; Ho, 2002; Sfekas, 2009). One exception is Avdic et al. (2019), who find a positive effect of hospital operation volume on patients' survival using Swedish register data on advanced cancer surgery procedures. They exploit the closures and openings of entire cancer clinics as an exogenous variation for volume in an instrumental variable set up. Importantly, they provide suggestive evidence that the effect on outcome is mainly due to increases in individual surgeon's experience.[7]

However, the literature testing for volume-outcome effects using individual (surgeon) level data is much more limited and it finds mixed results. 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. Using longitudinal data for a specific eye surgery (LASIK) in one clinic in Colombia, Contreras et al. (2011) find no effect of cumulative volume on outcome. Their set up benefits from a quasi-random allocation of surgeons to patients which makes selective allocation less of a problem. Similarly, Huesch (2009) fails to find any effect of cumulative volume on outcome for a panel of surgeons performing coronary artery bypass grafts (CABG) in Florida in the period 1998-2006. Moreover, he finds that almost all prior experience is depreciated from one quarter to the next. The author uses a choice model and predicted volume to mitigate potential issues of selective referral, although he does not reject exogeneity of volume.

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 -instead of cumulative volume. Ramanarayanan (2008) studies the same dataset for CABG surgeons as Huesch (2009) but uses the departure of a surgeon as an exogenous shock to the yearly volume of the remaining physicians. Instead, Huckman and Pisano (2006) do not discuss potential bias to a great extent and confine themselves to using surgeon 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.[8]

The current paper contributes to the existing literature in several ways. First, it provides new evidence of the causal link between patient outcomes and surgeon experience. As clearly showed before, the literature on volume-outcome at the individual level is in its early steps and

[7] Although they don't have data on surgeons, Avdic et al. (2019) suggest this is the main mechanism through a process of eliminating other possible alternatives.

[8] A related subset of papers find evidence of substantive human capital depreciation (forgetting) in the medical sector (Hockenberry and Helmchen, 2014; Huesch, 2009; Gaynor et al., 2005; David and Brachet, 2011; Gowrisankaran et al., 2006).

more research is needed. Second, previous studies looking at the causal effects of volume on outcome rely mostly on instrumental variable estimates. This paper serves as a complement to previous studies by exploiting a set up were selective referral is not possible, together with a dataset that allows to estimate the effect from within surgeon variation in volume. In addition, most previous studies use health care data from the United States, and focus almost exclusively on coronary artery disease procedures. Finally, the data employed allows me to make more precise estimates about the volume of patients seen by each surgeon in each point in time. Previous studies have relied mainly on yearly or quarterly data.

2 Clinical and Institutional Setting

2.1 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.

Based on their urgency degree, C-sections are typically classified in two groups: elective (or scheduled), and emergent (Lucas et al., 2000). The first group includes all C-sections scheduled in advance to occur before labor begins on the basis of an obstetrical or medical indication -although there is no immediate maternal or fetal compromise. The second group of C-sections includes all cases where the patient attempts to have a vaginal delivery (either through the natural onset of labor or medical induction) but end up delivering by C-section instead. This occurs when the patient develops 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.

2.2 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, in 2000 the Italian system provided the second best overall health care in the world -the first one being France (WHO, 2000). Furthermore it has the lowest maternal mortality rate worldwide at 1.94 for every 100,000 births (WHO et al., 2019).

Under this system, a pregnant woman cannot choose the physician or midwife that will assist her for the delivery unless she pays. Furthermore, given the well functioning of the

system, the grand majority (89%) of women choose to use the public service (Ministero della Salute, 2019).[9] This institutional feature eliminates the risk of selective referral, where institutions or surgeons with better performance attract higher volumes of patients -a common endogeneity issue in studies of learning-by-doing.

A typical patient in this system is assigned to the professional available at the time of admission to the hospital. Patients attempting labor are assisted by a midwife. If everything goes well and there are no complications, the midwife will help the patient through the entire process. However, if any complication arises, the midwife calls the physician in charge who will then evaluate whether a C-section is necessary.

In the year 2016, Italy had an overall C-section rate of 36.8%, with great disparities between the public (31.7%) and the private sector (50.9%) (Ministero della Salute, 2019).[10] It is worth mentioning that staff working in the delivery room in public hospitals are paid a fixed salary, meaning they have no personal financial incentive to recommend any particular treatment.

3 Empirical methodology

3.1 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:

$$\mathrm{y}_{ist} = \alpha + \beta\,\mathrm{e}_{st} + \delta\,\mathrm{d}_{st} + \mathbf{x}_{it}^{'}\,\theta + \phi_t + \eta_s + \epsilon_{ist} \tag{1}$$

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 four weeks leading up to and including the procedure on the patient surgeon s operated on just before operating on patient i.[11] d is a control for the number of days since the prior cesarean section surgeon s performed.[12] $\mathbf{x}_{it}$ contains individual-level control variables for

[9] For the data in hand, only a few dozen cases opted to pay, and are consequentially dropped from the study sample.

[10] According to OECD Health Statistics 2019, Italy has the seventh largest C-section rate among OECD countries.

[11] This measure is more precise than fixed calendar year as it responds instantaneously to any changes in the recent experience profile. I test for different time windows for recent experience below, from 4 to 52 weeks, and find complete forgetting beyond the last four weeks.

[12] In a recent study, (Hockenberry and Helmchen, 2014) show that a surgeon's number of days since last CABG is positively associated with patients' mortality rate. By construction, temporal distance and surgeon recent volume will be correlated. The correlation between these two is -.41 in my sample. I will show models

mother and pregnancy characteristics.[13] ϕ is a vector of indicators for year, month and day of the week of delivery.

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. Surgeon fixed effects ensures that the recent experience parameter in 1 is identified from changes in volume *within* surgeon. As discussed above, if physicians skills improve with recent 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.

One important assumption for the previous model to obtain causal effects is the lack of any compositional effect of patients between physicians with different levels of recent experience. To test whether selection bias affects my estimates, I regress each pre-treatment characteristic on the treatment -the number of C-sections performed in the last four weeks. If observed characteristics were associated with recent experience, it would be a sign of patient selection. The results for these estimations are reported in Figure 1 and Table A.1. After controlling for physician and time fixed effects, the treatment does not predict any of the observed maternal and pregnancy characteristics. This provides further evidence that mothers undergoing surgery with a physician with higher or lower recent experience are similar in observable characteristics.

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, hospitals aware of depreciating skills may assign healthier patients to physicians coming back from a period of low activity.[14] In this case my estimates on the impact of recent experience on patients' health would suffer from a downward bias.[15] To mitigate this concerns, I perform a series of robustness checks.

First, I use a restricted sample of more 'active' surgeons. Specifically, I repeat the analysis keeping only those surgeons that have performed at least 1 C-section in the four weeks before. This will help mitigate the possibility that my estimates of the effect of recent experience on outcomes are capturing some systematic unobservable quality differences between high and

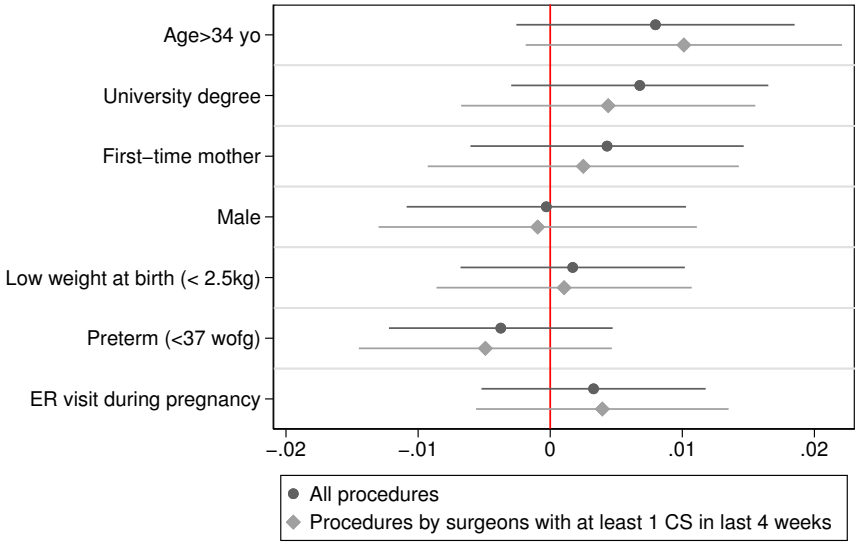
with and without temporal distance, but results are nearly identical.

[13]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 quadratic term for the number of gestational weeks, 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.

[14]This is what(Huesch and Sakakibara, 2009) call 'selective allocation'.

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

Figure 1: Balanced pre-treatment characteristics



Notes: The figure represents the coefficients and 95% confidence intervals from separate regressions of each predetermined variable on recent experience, controlling for days since prior C-section, physician fixed effects, and day-of-the-week, month, and year of birth fixed effects. Results for both the whole sample and the restricted sample of surgeons with at least 1 C-section in the last four weeks are presented.

low volume surgeons and patients' health.[16]

In a second robustness check, I estimate a separate coefficient for different types of C-sections depending on their emergency status. For patients admitted emergently, given the unexpected nature of these cases, surgeons need to make fast decisions under pressure and skill depreciation should be of particular relevance. For these patients undertaking an unplanned C-section, one important factor for the success of the surgery is the timing of the cut in relation to the contractions -since most of these patients are already in active labor at the time of the surgery.

3.2 Data

This study utilizes birth certificates from the maternity ward of a large public university hospital in Tuscany (Italy) for the years 2011 through 2014. The hospital has an average C-section rate of 31% average, close to the national Italian rate of 33% among public hospitals

[16]Furhtermore, the distribution of days since last C-section for the whole sample is highly positively skewed, with some surgeons showing gaps of more than 100 days (see Figure 4a in Appendix A). This is probably a result of doctors spending a period of time in other institutions that I can not observe in my data. However it would be wrong to assume that surgeons with gaps longer than 100 have lower recent experience.

in 2012.[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, age, previous deliveries, etc.), pregnancy characteristics (e.g. weeks of gestation, controls, assisted reproduction, etc.), birth characteristics (e.g. time of birth, type of labor, attendant, place, etc.) and indicators on newborns' health (e.g. weight, height, Apgar score, death, etc.). This information is complemented with surgeon's ID.[18]

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). From the 4,413 cases, I keep only one observation per pregnancy and drop 427 observations from plural births. In addition, I dropped from the analysis 86 births that had missing information in at least one of the variables used. Then I restrict the sample to surgeons who have performed, on average, at least 12 C-sections a year. This leaves the sample with 66 surgeons who performed 3,599 (92%) surgeries. Finally, I drop the first 30 days since the first surgery observed for each surgeon since I cannot compute experience for those. The final study sample has 3,468 births performed by 66 surgeons in the 4 years. As a robustness check, I will repeat the analysis for a sample of 'active' surgeons, namely those who have performed at least one C-section in the four weeks before the current surgery. This secondary study sample has 2,982 births performed by 59 surgeons.

Table 3.1 summarizes the variables used in the analysis. Averages for all variables except provider characteristics are virtually the same for both the whole and restricted samples. Mean admission to NICU was approximately 21%[19] and mean low Apgar score was 11%. As expected, emergently admitted patients have a higher probability of both having a low Apgar score and of being admitted to NICU than elective patients. The average age of patients is 34.5, and about 41% of them are first-time mothers -although this number is higher (49%) for non-elective procedures. About 21% of all births are born with less than 37 weeks of gestation.

[17] Hospital and national statistics were obtained from ARS Toscana (2014) and Ministero della Salut (2012).

[18] 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.

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

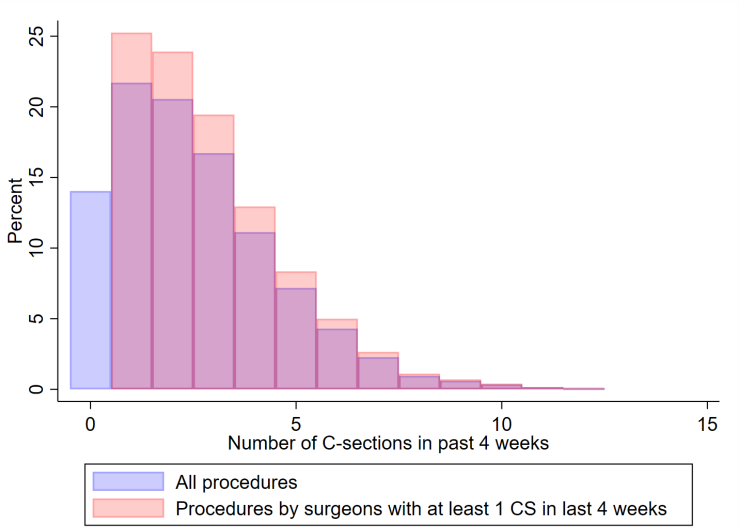
Table 3.1: Summary statistics

	All procedures				Procedures by surgeons with at least 1 CS in last 4 weeks			
	Any CS		Non-Elective CS		Any CS		Non-Elective CS	
	mean	sd	mean	sd	mean	sd	mean	sd
Outcomes								
% NICU	20.8	40.6	23.5	42.4	20.9	40.7	23.3	42.3
% Apgar score<9	11.0	31.3	14.0	34.7	11.3	31.7	14.4	35.1
Provider characteristics								
(mean) CS in past 4 weeks	2.5	2.0	2.8	2.0	3.0	1.9	3.2	1.9
(mean) days since last CS	14.8	24.4	12.0	18.2	8.4	7.9	7.6	7.6
Patient characteristics								
(mean) age	34.5	5.5	33.9	5.5	34.5	5.5	33.9	5.5
% university degree	30.5	46.0	29.3	45.5	31.2	46.3	29.9	45.8
% first-time mothers	40.5	49.1	48.7	50.0	41.0	49.2	49.0	50.0
Pregnancy characteristics								
% male	52.5	49.9	52.7	49.9	51.8	50.0	52.4	50.0
(mean) weight in grams	2,992.5	743.2	2,992.7	812.1	2,992.5	748.9	2,996.5	818.9
% low birthweight (< 2,500gr)	20.5	40.4	21.1	40.8	20.7	40.5	21.2	40.9
(mean) weeks of gestation	37.8	3.0	37.8	3.5	37.8	3.0	37.8	3.5
% preterm (<37 wofg)	21.0	40.7	22.0	41.5	20.6	40.5	21.6	41.1
% with at-least 1 ER visit	19.9	40.0	17.7	38.1	19.5	39.6	17.3	37.8
Observations	3599		1948		3048		1715	

Notes: Table contains variables used in the empirical analysis for the main estimation sample and the restricted sample of physicians who performed at least one CS in the past four weeks for the period 2011-2014. Statistics for the all CS and only non-elective CS are reported.

The mean number of procedures performed in the previous four weeks was 2.5 for all surgeons, and 3 for 'active' surgeons. Surgeons performing non-elective C-sections have a slightly higher mean recent experience than those performing elective procedures. Figure 2 shows the distribution of the measure of recent experience for both the whole sample and the restricted one.

Figure 2: Distribution of CS by surgeon's recent experience



Notes: The figure represents the distribution of recent experience for both the whole sample and the restricted sample of surgeons with at least one C-section in the past four weeks.

3.3 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.

4 Results

4.1 The effect of recent practice on patient health

Table 4.1 presents the average marginal effects of recent experience in a linear probability model.[20] Panel (A) shows coefficients for experience using the whole sample, while Panel (B) uses the sample restricted to physicians who performed at least one C-section in the four weeks prior to current ('active' surgeons). For each outcome, column one shows results of a model with controls, and physician and additive time fixed effects. Column two adds a control for surgeon's number of days since last C-section, column three uses clustered standard errors and column four adds surgeon's specific time-trends.

[20]Results using a probit model instead are virtually the same up to the second decimal.

Table 4.1: Effect of recent experience on birth outcomes

	Neonatal ICU				Apgar < 9			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (A): All procedures								
Experience (4w)	-0.006*	-0.008***	-0.008***	-0.007**	-0.001	-0.002	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Days since last CS		-0.001**	-0.001**	-0.001**		-0.000	-0.000*	-0.000**
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Observations	3,468	3,466	3,466	3,466	3,468	3,466	3,466	3,466
Mean dep.	0.208	0.208	0.208	0.208	0.110	0.110	0.110	0.110
Panel (B): Procedures by surgeons with at least 1 CS in last 4 weeks								
Experience (4w)	-0.008**	-0.010***	-0.010***	-0.008**	-0.005	-0.005	-0.005	-0.005
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Days since last CS		-0.002**	-0.002**	-0.001*		-0.000	-0.000	-0.001
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
Observations	2,982	2,982	2,982	2,982	2,982	2,982	2,982	2,982
Mean dep.	0.210	0.210	0.210	0.210	0.114	0.114	0.114	0.114
Cluster s.e. by surgeon			x	x			x	x
Surgeon trends				x				x

Notes: All models contain fixed effects for year, month and day of the week of birth, mother and pregnancy controls mentioned in Section 3.1 and surgeon fixed effects. NICU stands for Neonatal Intensive Care Unit. Standard errors are in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Panel (A) shows that an increase in recent experience was associated with a decrease in the likelihood of NICU admission. Specifically, one additional C-section in the previous four weeks reduces the probability of being transferred to NICU by about 0.7 percentage points. To put the estimate in context, this result implies that a one standard deviation increase in the average number of C-section performed in the last four weeks, an increase of two C-sections, is associated with a 1.4 percentage-point reduction in need for NICU, or 6.7% of the sample mean use of NICU. The estimates are consistent and stable across specifications. I fail to find any significant effect on the likelihood of having an Apgar socre below 9. Results are qualitatively the same when restricting the sample to 'active' surgeons (Panel B), although coefficients are of a larger magnitude. Finally, estimates of the association between days since last C-section and outcome, although some times statistically different from zero, are quantitatively very close to zero.[21]

Table 4.2 reports estimates by type of C-section, with the first two columns reporting estimates for elective births and the next two for non-elective cases. Again, Panel (A) reports estimates for the whole sample while Panel (B) uses the sample restricted to 'active' surgeons. The first thing to notice is that the effect of recent experience on outcome is present only for non-elective procedures. Estimates for elective surgeries show precisely estimated zero effects of experience on outcome. On the other hand, the effect on emergency procedures are about double the size of the whole sample estimates from previous table. Specifically, a one standard deviation increase in the average number of C-sections performed in the previous four weeks, an increase of two C-sections 1.9 C-sections, is associated with a 2.9 percentage-point reduction in need for NICU, or 12% of the sample mean use of NICU. Furthermore, the effect on the likelihood of a low Apgar score are also statistically significant. A one standard deviation increase in recent experience implies a 1.3 percentage-point reduction in likelihood of having an Apgar score below 9, or 9.5% of the sample mean. Results for the restricted sample are similar in magnitude.

4.2 Different time-windows of recent experience

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. To test how sensitive results are to this decision, I run a set of regression for different time spams (from

[21]Furthermore, in auxiliary regressions where recent experience is not included the estimates are not statistically significant.

Table 4.2: Effect by type of C-section

	Elective		Non-Elective	
	NICU	Apgar<9	NICU	Apgar<9
Panel (A): All procedures				
Experience (4w)	0.000	0.001	-0.015***	-0.007*
	(0.004)	(0.004)	(0.003)	(0.004)
Days since last CS	-0.000	-0.000	-0.000	-0.001***
	(0.000)	(0.000)	(0.001)	(0.000)
Observations	1,576	1,576	1,889	1,889
Mean dep.	0.176	0.076	0.234	0.139
Panel (B): Procedures by surgeons with at least 1 CS in last 4 weeks				
Experience (4w)	-0.001	0.000	-0.015***	-0.009**
	(0.005)	(0.005)	(0.003)	(0.005)
Days since last CS	-0.002*	0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1,297	1,297	1,681	1,681
Mean dep.	0.180	0.075	0.234	0.144

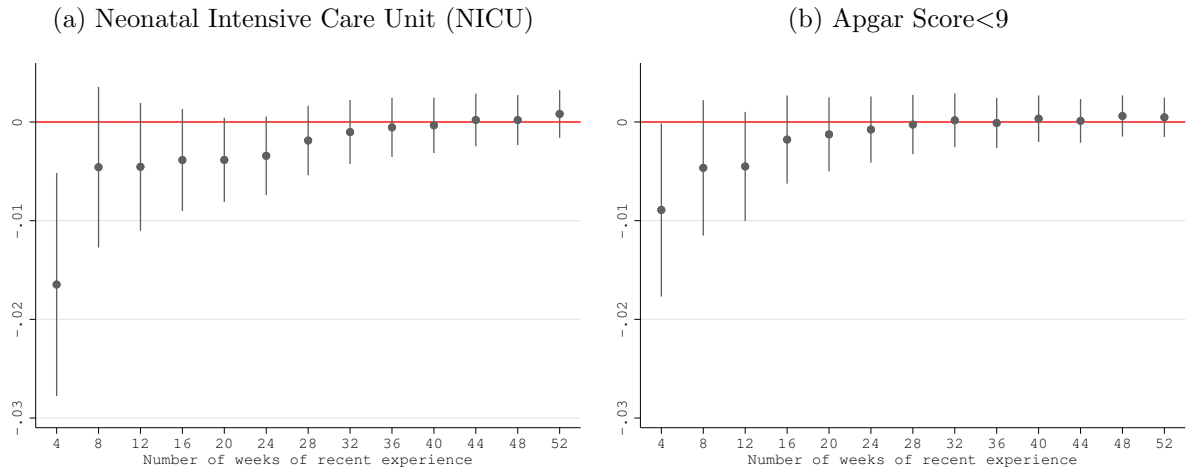
Notes: Sample of cases with a surgeon who has done at least one CS in the last four weeks. All models contain fixed effects for year, month and day of the week of birth, mother and pregnancy controls mentioned in Section 3.1 and surgeon fixed effects. ICU stands for Intensive Care Unit. Standard errors clustered by surgeon are in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$

4 weeks up to 52 weeks) for the two outcomes. Figure 3 shows the results for the case of non-elective 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. This provides further evidence for the human capital depreciation hypothesis, where procedures performed further back in time have little effect on surgeon's ability today -controlling for her average ability.

5 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 feature of the Italian health care system -patients are not allowed

Figure 3: Effect of recent experience using different time windows



Notes: Each black dot is the average marginal effect of recent experience using the time window specified in the x-axis on the probability of needing NICU or having an Apgar Score below 9 (controlling for year, month and day-of-the-week of birth, and surgeon fixed effects). 90% confidence bounds indicated by the vertical grey lines. In order to keep the same sample given that I use a lagged independent variable, all regressions exclude data for the year 2011.

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, my empirical approach rests on an institutional context that allows me to estimate parameters that are free from selective referral bias. Second, I provide 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 between 2011 and 2014, I find that, for emergent cases, a one standard deviation increase in recent experience reduces the likelihood of neonatal intensive care unit admission by nearly 2.9 percentage points (12%) and of being born with a low Apgar Score by about 1.3 percentage points (9.5%), all else equal. This effect is not present for the case of elective C-sections.

The results of this study would suggest that learning-from-recent-experience effects may be substantive. These findings support recent initiatives to favor higher volume providers when making volume allocation decisions.

References

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

Avdic, D., Lundborg, P. and Vikström, J. (2019), 'Estimating returns to hospital volume: Evidence from advanced cancer surgery', *Journal of health economics* **63**, 81–99.

Birkmeyer, J. D., Stukel, T. A., Siewers, A. E., Goodney, P. P., Wennberg, D. E. and Lucas, F. L. (2003), 'Surgeon volume and operative mortality in the united states', *New England Journal of Medicine* **349**(22), 2117–2127.

Chowdhury, M., Dagash, H. and Pierro, A. (2007), 'A systematic review of the impact of volume of surgery and specialization on patient outcome', *British Journal of Surgery: Incorporating European Journal of Surgery and Swiss Surgery* **94**(2), 145–161.

Contreras, J. M., Kim, B. and Tristao, I. M. (2011), 'Does doctors' experience matter in lasik surgeries?', *Health economics* **20**(6), 699–722.

David, G. and Brachet, T. (2011), 'On the determinants of organizational forgetting', *American Economic Journal: Microeconomics* **3**(3), 100–123.

Gaynor, M., Seider, H. and Vogt, W. B. (2005), 'The volume-outcome effect, scale economies, and learning-by-doing', *American Economic Review* pp. 243–247.

Gowrisankaran, G., Ho, V. and Town, R. J. (2006), 'Causality, learning and forgetting in surgery', *Mimeo* .

Halm, E. A., Lee, C. and Chassin, M. R. (2002), '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.

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

Ho, V. (2014), Learning by doing, *in* A. J. Culyer, ed., 'Encyclopedia of Health Economics', Elsevier, San Diego, pp. 141 – 145.
URL: *http://www.sciencedirect.com/science/article/pii/B978012375678701110X*

Hockenberry, J. M. and Helmchen, L. A. (2014), 'The nature of surgeon human capital depreciation', *Journal of health economics* **37**, 70–80.

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

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

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

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

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

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

Ministero della Salut (2012), Certificato di assistenza al parto (cedap). analisi dell'evento nascita - anno 2012, Technical report.
URL: *http://www.salute.gov.it/imgs/C_17_pubblicazioni_2768_allegato.pdf*

Ministero della Salute (2019), Certificato di assistenza al parto (cedap): Analisi dell'evento nascita - anno 2016, Technical report.
URL: *http://www.salute.gov.it/imgs/C_17_pubblicazioni_2881_allegato.pdf*

Osservatorio di Epidemiologia dell'Agenzia Regionale di Sanità (ARS) della Toscana (2014), Nascere in toscana 2012, Technical report.
URL: *https://www.ars.toscana.it/files/pubblicazioni/in_cifre/2_Incifre_CAP_web.pdf*

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

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

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

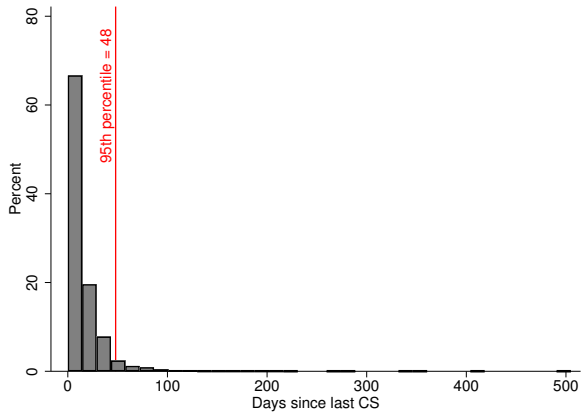
WHO (2000), *The world health report 2000: health systems: improving performance*, World Health Organization.

WHO, UNICEF, UNFPA, WBG and UNDP (2019), Trends in maternal mortality: 2000 to 2017: Estimates by WHO, UNICEF, UNFPA, World Bank Group, and United Nations Population Division., Technical report.
URL: *https://www.unfpa.org/sites/default/files/pub-pdf/Maternal_mortality_report.pdf*

Appendix A Other graphs and tables

Figure 4: Distribution of CS by surgeon's days since last CS

(a) All procedures



(b) Procedures by surgeons with at least 1 CS in last 4 weeks

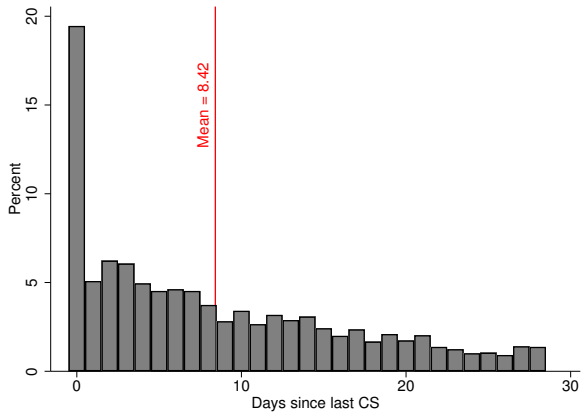


Table A.1: Balanced pre-treatment characteristics

	Mother's Age>34	University degree	First-time Mother	Male Newborn	Low weight at birth	Preterm birth	Emergency visit
Panel (A): All procedures							
Experience (4w)	0.008	0.007	0.004	-0.000	0.002	-0.004	0.003
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Observations	3,466	3,466	3,466	3,466	3,466	3,466	3,466
Panel (B): Procedures by surgeons with at least 1 CS in last 4 weeks							
Experience (4w)	0.010*	0.004	0.003	-0.001	0.001	-0.005	0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Observations	2,982	2,982	2,982	2,982	2,982	2,982	2,982

Notes: Table contains the coefficients and standard errors from separate regressions of each predetermined variable on the treatment (experience), controlling for surgeon, day-of-the-week, month, and year of birth fixed effects. Panel (A) shows estimates using the whole sample, while Panel (B) uses the sample restricted to physicians who performed at least one C-section in the four weeks prior to current ('active' surgeons).

Últims documents de treball publicats

Núm	Títol	Autor	Data
20.09	Low Staffing in the Maternity Word: Keep Calm and call the Surgeon	Gabriel A. Facchini Palma	Setembre 2020
20.08	Impact of public transport strikes on air pollution and transport modes substitution in Barcelona	Lyna González / Jordi Perdiguero / Àlex González	Setembre 2020
20.07	Globalización y responsabilidad en los problemas ecológicos	Jordi Roca Jusmet / Emilio Padilla Rosa	Setembre 2020
20.06	Do short-term rental platforms affect housing markets? Evidence form Airbnb in Barcelona	Miquel-Àgel Garcia / Jordi Jofre / Rodrigo Martínez / Mariona Segú	Setembre 2020
20.05	The driving factors of CO2 emissions from electricity generation in Spain: A decomposition analysis	Vicent Alcántara / Emilio Padilla / Pablo del Río	Juliol 2020
20.04	CO_2 What do divided cities have in common? Am international comparison of income segregation	Paolo Veneri/ Andre Comandon / Miquel-Àngel Garcia López / Michiel N.Daams	Juliol 2020
20.03	CO_2 emissions of the construction sector in Spain during the real estate boom: input-output subsystem analysis and decomposition	Vicent Alcántara/ Emilio Padilla	Juny 2020
20.02	The Direct Rebound Effect of Electricity Energy Services in Spanish Households: Evidence from Error Correction Model and System GMM estimates	Martín Bordón / Jaume Freire / Emilio Padilla	Maig 2020
20.01	Subsidizing Innovation Over the Business Cycle	Isabel Busom / Jorge-Andrés Vélez-Ospina	Març 2020
19.07	África Subsahariana: ¿Del afropesimismo a la transformación económica?	Artur Colom Jaén	Novembre 2019
19.06	Identification of relevant sectors in CO2 emissions in Ecuador through input-output analysis	Edwin Buenaño / Emilio Padilla and Vicent Alcántara	Setembre 2019
19.05	Driving forces of CO2 emissions and energy intensity in Colombia	Lourdes Isabel Patiño / Vicent Alcàntara and Emilio Padilla	Setembre 2019
19.04	The relation of GDP per capita with energy and CO2 emissions in Colombia	Lourdes Isabel Patiño / Vicent Alcàntara and Emilio Padilla and Josep Lluís Raymond	Setembre 2019
19.03	Cruise activity and pollution: the case of Barcelona	Jordi Perdiguero / Alex Sanz	Juliol 2019