Understanding productivity in maternity wards

Marina di Giacomo*, Massimiliano Piacenza**, Luca Salmasi***, Gilberto Turati***

* Department of Economics, Social Studies, Applied Mathematics and Statistics - University of Torino

** Department of Economics and Business University of Piemonte Orientale (Novara)

*** Department of Economics and Finance, Catholic University, Rome

Corresponding author: Marina Di Giacomo, Department of Economics, Social Studies,

Applied Mathematics and Statistics - University of Torino. E-mail: marina.digiacomo@unito.it

Very Preliminary Draft

Abstract

In this paper we estimate health personnel productivity in causal terms. We exploit a discontinuity in health personnel presence induced by law. We argue that hospitals above the 500 births threshold have incentives to hire more health personnel to comply with law requirements in order to obtain accreditation, even though they have very similar workloads with respect to hospitals located just below the threshold. We exploit this discontinuity to estimate an instrumental variable model where the probability of having a full team (i.e., gynecologist, midwife and pediatrician) during delivery is affected by having a number of births above 500, and we use this exogenous variation to estimate the effect on newborns' and maternal health status. We find that having number of births above the threshold increases the probability of full team by almost 20% points and this has significant effects on health. Heterogeneous effects are investigated across day of week, time of day and congestion level.

Keywords: Maternity wards productivity; Instrumental variables; Regression Discontinuity Design

JEL classification: I10; I12.

1 Introduction

Health care expenditure has been rapidly growing in the past decades, and available estimates suggest that it will experience further increase in the future almost everywhere in the world. There are different views of why health expenditure is growing. Demand side explanations based on an aging population have been largely disputed (Zweifel et al., 1999; Seshamani and Gray, 2004). Supply side explanations are divided between the classical Baumol's disease, and modern views suggesting that health care is not different from any other industries with respect to productivity. According to this latter view, health expenditure is increasing because health care is plagued by large inefficiencies and waste. Over-treatments, adoption of more costly and less effective technologies, inappropriateness are all examples showing that there is room for improvement. Reducing inefficiency and waste, and enhancing productivity, may be a way to cope with a rising demand for health care services and the austerity of budgets.

Unsurprisingly, given its importance, the issue of measuring and explaining productivity in health care has gained attention among scholars (Chandra & Staiger 2007, Chandra et al. 2010, Chandra et al. 2016, Bloom et al. 2014). In this paper we assess the productivity of maternity wards. We do so by documenting the effect of medical staff availability in delivery rooms on newborns' health, exploiting the exogenous variation in personnel requirements defined on the number of birth deliveries by the Italian legislation. The thresholds defined by the law allow us to identify the causal impact of increasing medical staff on hospital productivity, avoiding the simultaneity bias deriving from the choice of staff given the needs of patients.

The reorganization of maternity wards in Italy started in 2010. The goal of the reform was to maintain only high-volume maternity wards, given the widely accepted evidence of a positive association between volumes and health outcomes (Gaynor et al. 2005, Snowden et al. 2012, Giancotti et al. 2017). Maternal wards treating less than 500 births per year should have been progressively closed. Remaining maternal units were then sorted into two levels by the law: 'first level' maternity wards are those treating at least 500 births per year, while 'second level' maternity wards those treating more than 500 births per year. The reform also encompasses a precise definition of the number of medical staff units (midwives, gynecologists and pediatricians) per work shift for both first and second level units.

The reorganization was slowly implemented and among the 561 maternity wards operating in Italy in 2011, 192 had less than 500 births per year, and only 88 were closed by 2016. Technical reports by National Health authorities highlight the many difficulties encountered during the reorganization, mainly related to the lack of a shared method to compute hospitals' health personnel requirements, the lack of resources to correct inappropriate situations, and the fierce opposition of communities subject to the closure of their maternity ward.

Our analysis is based on a RDD framework focusing on maternity wards around the 500 births per year threshold in the first period after the reorganization (2011-2013). Our productivity measure is based on the health of newborns, controlling for a large set of mother characteristics which account for differences in needs during delivery. In order to identify the effect of health personnel availability on newborns health, while accounting for potential endogeneity of the medical staff, we exploit the discontinuity induced by having a number of births above the 500 births threshold as instrument for the availability of health personnel. We estimate Local Average Treatment Effects (LATE) considering different health outcomes: (i) the Apgar scores at 1 and at 5 minutes, i.e. two indexes (in the 1-10 range) about the newborn general health status measured one minute and five minutes after birth ; (ii) a binary indicator for the need of infant resuscitation.

First, the law appear to influence medical staff availability. Considering natural deliveries, we find that the probability of having a full team (i.e. gynecologist, midwife and pediatrician) in the delivery room increases by almost 12 percentage points for hospitals located above the 500 births threshold. We obtain even larger increases for scheduled C-sections, for which the probability of having a full team increases by almost 21 percentage points. More important, our findings suggest that a full team (as with respect to a smaller team) has a sizable and significant effect on the health of newborns. We find that a 10 percentage points increase in the presence of a full team of specialists: (i) decreases the probability of the newborn resuscitation by almost 0.7 percentage points; (ii) increases the Apgar score after 1 minute by 0.076 points; (iii) increases the probability of having a perfect Apgar score after 5 minutes by 0.87 percentage points.

The remainder of the paper is organized as follows: Section 2 describes the institutional setting, Section 3 introduces the data and the descriptive evidence, Section 4 discusses the empirical strategy and the results, while Section 5 concludes.

2 Institutional background

The Ministerial Decree (MD) 24/04/2000 defines how obstetric and pediatric or neonatal services are organized at the regional, or inter-regional, level in the Italian Health Service (IHS). The decree distinguishes between obstetric and neonatal intensive care units, with the latter entitled to treat newborns requiring intensive care, and defines for each type of unit three levels of assistance according to: hospitals' number of births per year and mothers' gestational age at delivery. [1] According to this decree we can define: (i) first level maternity wards, which should treat at least 500 births per year and women with gestational age smaller than 34 weeks; (ii) second level wards, which should treat at least 800 births per year and women with gestational age at birth greater than 32 weeks, and (iii) third level wards, which should treat at least 1,000 births per year and all women, irrespective of their gestational age. Each level has to meet a number of standards of quality regarding the constant presence of obstetric and gynecologic personnel and a sufficient number of beds and delivery rooms. Moreover, each unit must guarantee a 24/7 presence of gynecologic, anesthesiology, and pediatric personnel with specific competences for neonatal resuscitation and the presence of an obstetric coordinator, midwives, pediatric nurses and/or social-health operators.

After the 5-th December 2010 the State-Regions conference introduced some novelties in the legislation of maternity wards. First, the number of levels was reduced from 3 to 2, and second and third level wards were merged in a unique category. Second, the ideal number of midwives per shift was defined more precisely. As shown in Table 1 the recommendation for first and second level wards was to have at least 2, 3, 4 or 5 midwives per shift if their annual number of births was between 500 and 999, 1,000 and 1,499, 1,500 and 1,999, respectively. Above 2,000 births per year an extra midwife per shift is added every 750 extra births per year. This legislation can be modified by Regions that can decide autonomously how to organize obstetric and gynecologic services within the limits set by national guidelines.

– Insert Table 1 here –

[1] Gestational age is the number of weeks between conception date and delivery date. A preterm (or premature) birth occurs before the 37th pregnancy week. A term birth is between 37th and 41st week. A postterm (prolonged) pregnancy occurs when delivery is on or after 41 pregnancy weeks

In order to be accredited as first or second level wards, hospitals must meet all the operative, security and technological standards defined by Ministero della salute (2010). Each ward is evaluated by a National Birth Path Committee (NBPC) and particular attention is given to wards below the 500 births threshold, that may be allowed to operate if they serve geographical areas with accessibility problems. In this case regional health departments have to provide a motivated request to the NBPC that evaluates if the ward meets the necessary quality and security requirements, and if not, defines a program to overcome the problems found. The NBPC has 3 months to formulate a judgment and, if adjustments are required, the maternity ward has 3 months to comply. After this procedure the ward is constantly monitored by the regional health department and by the NBPC, that verify its performance in terms of maternal and newborns' health status.

In this paper we use data from maternity wards in Piedmont, where decisions about the organization of health services are described in the regional social-health plan. The plan is discussed every three years and defines regional perspectives on various topics related to health services. As regards maternity wards one of the main aspects debated is related to the progressive closure of wards with less than 500 births per year (Regione Piemonte 2012). The objective is to maintain in the long run only wards granting at least 1,000 births per year, since high-volume hospitals typically perform better than low-volume ones, allowing for some exceptions only to satisfy the needs of particular areas that for geographic reasons cannot reach high volumes of births.

The reorganization of maternity wards was proposed in 2010 by the NBPC and was supported by the Ministry of Health. However, the situation is not yet resolved and all regions still have some wards who do not meet the 500 births per year criterion. In fact, among the 561 maternity wards operating in Italy in 2011, 192 had less than 500 births per year and only 88 were closed by 2016. Among those still open the majority are located in southern Italy (53 wards). Piedmont had 9 out of 34 wards below the 500 births threshold in 2011, 4 of which were closed before 2016, La Stampa (2016). Despite all these procedures to keep high quality standards a report performed by the Regione Piemonte (2012) highlighted how there are still many difficulties encountered during accreditation which are mainly related to the lack of a shared scientific method to calculate hospitals' health personnel requirements, and of resources to correct inappropriate situations. Thus, especially wards below the threshold may have difficulties in meeting the standards recommended by law, and related to health personnel availability. In addition, from the 29-th July of 2010 to the 21-st March of 2017 Piedmont was subject to a recovery plan, adopted to control regional health care expenditure, that strongly limited the possibility to hire medical personnel beyond turn over. We believe that the actual legislation does not account for the fact that defining sharp thresholds may provide an incentive to invest in health personnel only to hospitals above thresholds even though those right below may have very similar characteristics in terms of workloads. This mechanism could favor inequalities of opportunity for women who use maternity services.

Unfortunately, we do not have direct information about health personnel variations in the years of our analysis but we use external data on public competitions to describe how jobs as gynecologist, midwife and pediatrician varied during the years 2008-2016 in maternity wards located in Piedmont, see Table 2. The Table shows the number of gynecologists, pediatricians and midwives hired by time period and volume of births. Unsurprisingly the majority of open public competitions for personnel recruitment took place before 2011, years in which the recovery plan did not impose restrictions in terms of personnel hirings in Piedmont. It is clear how, especially during the years of the recovery plan health personnel was allocated in greater numbers especially to hospitals above the 500 births threshold.

– Insert Table 2 here –

Also in relation to the problems highlighted above Piedmont started in 2016 a process to formally determine hospitals' health personnel requirements, defining for maternity wards a rule that assigns an extra midwife every 125 deliveries per year, with a minimum of 6 midwives per ward (Regione Piemonte 2016). In order to realize the objectives proposed in these guidelines, and given the higher budget availability due to the end of the recovery plan, the Regione Piemonte (2017) provided funds to hire, beyond standard turn over, 1,800 workers in the health sector, especially nurses, midwives and social-health operators.

3 Data and methods

3.1 Data

Our study is based on data from the Certificate of Delivery Assistance (CeDAP) of the Piedmont region in Italy for the years 2002-2017. In Italy, state law requires certificates to be compiled for all births and each region is required to compile the same form in order to produce comparable indicators. The midwife who attends the birth, or the doctor responsible for delivery in the operating room, must complete the certificate within 10 days of delivery. The certificate lists epidemiological information regarding mothers' health status, socio-demographic characteristics, risk factors during pregnancy, obstetric procedures, characteristics and method of delivery, and includes any abnormal condition or congenital anomaly of the infant, cause of mortality, information about use of prenatal care services, etc. (for further details, see Decree No. 349 of the Italian Ministry of Health).

For the aims of our analysis we consider the following indicators of newborns' health status at birth: (i) the apgar score measured after 1 and 5 minutes[2] and (ii) the probability of no need of resuscitation. We decided to analyse these outcomes because they have been widely studied in the health economics literature, and in particular apgar scores have been found to be predictive of health, cognitive ability, and behavioral problems of children at age three (Almond et al. 2005), of reading and math test scores in grades 3-8 (Figlio et al. 2014), and of school attainment, earnings and social assistance receipt after age 18 (Black et al. 2007, Oreopoulos et al. 2008).

The apgar score can be classified as follows: scores of 7 and above are considered as normal, with 9 and 10 considered as a perfect score, from 4 to 6 are considered fairly low, and 3 and below are critically low. A low score on the 1 minute index indicates that the newborn requires medical attention but does not necessarily correlates with long-term health problems, especially if the score improves at the 5 minutes threshold.

Resuscitation is defined according to a variable that records the need to treat newborns with drugs, intubation, cardiac massage or oxygen at birth.

Table 3 lists descriptive statistics for the variables of interest in our study for hospitals with yearly number of births below 1,000 (all) and for hospitals with yearly number of births between 350 and 650 (discontinuity sample).

– Insert Table 3 here –

[2]The Apgar score is based on a score ranging from 0 to 10, assigned on the basis of 5 criteria: appearance, pulse, grimace, activity and respiration and is measured after 1 and 5 minutes from delivery.

3.2 Our sample

As already discussed in previous sections we limit our analysis to first level wards, since we believe that the lack of personnel may be more prominent below this threshold. We exclude deliveries before the 34-th gestational week - since they can not be treated by first level wards - urgent C-sections and ventouse/forceps deliveries, because of the small number of observations. We consider only the years between 2011 and 2013 because before this period wards were classified in three levels rather than two. We excluded information after 2013 because some wards were closed implying possible problems of attrition if only more productive wards survive.

For the aims of this analysis we considered only hospitals with an absolute distance between their 5-years moving average[3] and 500 not larger than 150. We chose this value to guarantee an adequate number of hospitals in the analysis, and to exclude hospitals with a large distance from 500 births. In fact, those hospitals with number of births higher (lower) than 500 and close to 1,000 (0) could have health personnel availability more similar to that of hospitals located right above (below) the 500 births threshold than of those located below (above), implying a possible downward (upward) bias in our estimates. Our final sample is composed by 8,743 deliveries.

3.3 The delivery path

In order to understand productivity of maternity wards it is useful to give a detailed description of the delivery path followed by pregnant women, and to describe the duties of health personnel involved in each phase of the process. For the aims of the paper we will focus only on the last part of delivery, which is composed by labour and birth. The

[3] More details about the construction of this indicator are provided in following sections

standard team present in the delivery room is generally composed by a gynecologist and/or a midwife, whereas a pediatrician/neonatologist may be present for deliveries with or at risk of complications. From our sample we can obtain the distribution of health personnel by type of delivery and number of births.

In Table 4 we show the number of births and the composition of health personnel in the delivery room by type of delivery and number of births for the years 2011-2013. We consider a total of 13,140 deliveries, with normal deliveries and scheduled C-sections accounting for 64% and 26% of the total. The remaining 10% is composed by urgent C-sections (6%) and ventouse/forceps deliveries (4%). As regards the distribution of health personnel we can see from Table 4 that the probability of having a full team (i.e., gynecologist, midwife and pediatrician) increases by almost 12 % points for hospitals located above the 500 births threshold. The same happens when we consider scheduled C-sections, where the probability of having a full team increases by almost 21 % points. Other types of delivery represent only 10% of total deliveries and for this reason won't be considered in the analysis.

– Insert Table 4 here –

3.4 Empirical strategy

In order to study how productivity varies across maternity wards we first propose the following model:

$$\begin{aligned} Y_{iht} = \lambda_h HH + \lambda_t T + \lambda_{h,t} HH \times T + \lambda_1 HP^{'} \boldsymbol{\lambda_{HP}} + \\ + M^{'} \boldsymbol{\lambda_M} + D^{'} \boldsymbol{\lambda_D} +_{it}, \end{aligned} \tag{1}$$

Y_{dht} represents the share of babies born in good health (i.e, with no necessity of resuscitation or with apgar score ≥ 7) discharged by hospital H during day d in year t. In order to measure total factor productivity (TFP) we follow Chandra et al. (2016) and estimate a model with a full set of interactions between hospitals and years $\lambda_{h,t}$ controlling for the presence of different teams of specialists during delivery (considered as hospital inputs - HP), and maternal M and delivery D characteristics.

Estimates of λ_{HP} are likely to be biased if hospitals choose different observable input levels for patients who differ unobservably in their latent survival, or if their choice of unobservable inputs is correlated with observed inputs.

In order to solve this problem we will refer to an IV specification where having number of births above 500 will be used as an exogenous variation of health personnel availability. Our preferred specification will then include a measure of the distance from the threshold of 500 births to account for hospital specialization and linear time trends as well as local health authorithy (lha) fixed effects.

We look at the effect of health personnel availability on newborns' health. Let us consider the following specification:

$$\begin{aligned} Nbh_{ihpt} = \alpha_{pt} + \alpha_1 hp^{full}_{ihpt} + \alpha_2 tr_{mt} + \alpha_3 tr^2_{mt} + \alpha_4 q_{ht} + \alpha_5 q^2_{ht} + \\ + M^{'}\boldsymbol{\alpha_M} + D^{'}\boldsymbol{\alpha_D} + H^{'}\boldsymbol{\alpha_H} + \epsilon_{it}, \end{aligned} \tag{2}$$

Nbh_{ihlt} measures health status at birth for newborn i in hospital h, local health authority (lha) p and time t. Health is measured through (i) the Apgar score at 1 and 5 minutes, considered as a continuous measure or as dichotomous indicator of the probability to have an Apgar score lower than a given value (7, 8 or 9) or as the probability of perfect score (9 or 10); (ii) an indicator for the need of resuscitation, measured by a variable that records whether right after delivery the newborn needed to be treated with drugs, intubation, cardiac massage or needed oxygen; hp^{full}_{ihpt} identifies the presence of deliveries where a full team of specialists (i.e., gynecologist, midwife and pediatrician) is available during delivery i in hospital h, lha p and time t. We include a quadratic time trend, tr_{mt} and tr^2_{mt} that varies by year t and month m, to capture outcome variations correlated with exogenous regional shocks. $M = \{M_k\}_{k=4}^{K_1}$ is a matrix of mother-specific characteristics, including information about mother's age, nationality, education, employment, marital status and previous deliveries. $D = \{D_k\}_{k=K_1}^{K_2}$ is a matrix of delivery-specific characteristics including information about delivery week and type, day of week, and month fixed effects. $H = \{H_k\}_{k=K_2}^{K_3}$ is a matrix of hospital-level characteristics including information about the presence of a pediatrician 24/7 or upon call in the hospital. Lastly, α_{pt} represent lha fixed-effects[4].

In this framework we are interested in α_1, the coefficient that captures the effect of hp^{full}_{ihpt} on Nbh_{ihpt}. In this setting the variable hp^{full}_{ihpt} captures the effect of an increase in the probability of having a pediatrician during normal deliveries, since the main reference category below the threshold is represented by gynecologist and midwife, and of an increase in the probability of having a midwife during scheduled C-sections, since the main reference category in this case is represented by gynecologist and pediatrician.

Standard ordinary least squares (OLS) estimates of α_1 yield unbiased results if the orthogonality condition is satisfied; i.e. $E\left[hp^{full}_{ihpt}\epsilon_{ihpt}\right] = 0$. However, this is unlikely

[4]We did not include hospital fixed-effects since in the period considered there is a very limited within variation across the 500 births threshold at the hospital level.

to hold because of the presence of endogeneity caused by nonrandom allocation of staff. Hospitals (and firms in general) take decisions about investments and optimal levels of input usage based on their level of productivity. A low-productivity level may represent an incentive to hire more personnel (i.e., reverse causality) also known as endogenous productivity in the R&D literature (see for instance Marschak & Andrews (1944)).

To account for endogeneity we use an instrumental variable approach adopting the following empirical specification:

$$\begin{aligned} Nbh_{ihpt} &= \gamma_0 + \gamma_1 hp^{full}_{ihpt} + \gamma_2 tr_{mt} + \gamma_3 tr^2_{mt} + \gamma_4 q_{ht} + \gamma_5 q^2_{ht} + \\ &+ M^{'}\boldsymbol{\gamma_M} + D^{'}\boldsymbol{\gamma_D} + H^{'}\boldsymbol{\gamma_H} + \epsilon_{it}, \\ hp^{full}_{it} &= \beta_0 + \beta_1 Above\_threshold_{ht} + \beta_2 tr_{mt} + \beta_3 tr^2_{mt} + \beta_4 q_{ht} + \beta_5 q^2_{ht} + \\ &+ M^{'}\boldsymbol{\beta_M} + D^{'}\boldsymbol{\beta_D} + H^{'}\boldsymbol{\beta_H} + \eta_{it}, \end{aligned} \tag{3}$$

where γ_1 represents the two-stages least squares (IV-TSLS) estimate of the parameter of interest.

Productivity is affected by health personnel availability and by their level of experience, that depend on hospitals' total number of births, for the influence of endogenous productivity and of learning by doing. In order to identify the effect of health personnel availability on newborns' health from that of experience, accounting for reverse causality, we employ a Regression Discontinuity Design (RDD) where the discontinuity induced by having number of births above the 500 births threshold is used as instrument for health personnel availability and allows us to estimate the relationship in causal terms and to obtain the Local Average Treatment Effect (LATE) (Imbens & Angrist 1994).

In order to control for health personnel experience we define the terms q and q^2 as follows: first we calculate a moving average of the number of births for hospital h in the period t, $t-5$ ($births_{h,(t,t-5)}$), second, we obtain the distance from the 500 births threshold as $dist_{ht} = births_{h,(t,t-5)} - 500$. Lastly, we define $q_{ht} = dist_{ht} + 150$, and q^2_{ht} as its squared value, so that q_{ht} takes only positive values.

We decided to use a 5-years moving average under the assumption that decisions about optimal levels of health personnel are taken considering medium term variations in number of births rather than (possibly) idiosyncratic short-term shocks[5].

$Above\_threshold_{ht}$ is our instrumental variable and is a dummy that takes value 1 if the distance between the 5-years moving average and the 500 births threshold is greater than 0 ($dist_{ht} > 0$). In this setting we assume that hospitals located right above or below the threshold are very similar to each other in terms of workload, medical equipment and infrastructural endowments, but those above have an incentive to hire more health personnel.

4 Results

4.1 Estimates of productivity levels

Figure 1 shows productivity dispersion estimated using the coefficients of hospital/years interactions from equation (1). As we can see there is a large variability in productivity levels across hospitals, but hospitals above the 500 births threshold show higher levels of productivity in terms of probability to do not need resuscitation and of having an healthy apgar (≥ 7).

– Insert Figure 1 here –

[5] As a robustness we used different time windows to define the moving average on number of births and found that results do not change. These Tables are available from the authors upon request.

4.2 Preliminary evidence

Indeed, descriptive evidences shown in previous sections should be formally tested and health personnel variations across thresholds should be evaluated netting out the effect of other potential confounders. For this reason we propose a graphical analysis, where we plot the residuals of the following regression:

$$Y_{it} = \beta_0 + \eta_1 tr_{mt} + \eta_2 tr_{mt}^2 + \eta_3 q_{ht} + \eta_4 q_{ht}^2 + \\ + M^{'} \boldsymbol{\eta_M} + D^{'} \boldsymbol{\eta_D} + H^{'} \boldsymbol{\eta_H} + \eta_{it},$$

Y_{it} represents the presence of one of the team of professionals listed in Table 2 during delivery i at time t, or one of the health outcomes already described in the previous Section. This specification is similar to that presented in the first stage regression of our IV model (equation 2), apart from the presence of the variable $Above\_threshold_{it}$. In this way we can see if a jump in treatment or outcome variables still persists after controlling for other potential confounders.

Figures 2 and 3 display the probability of having a given team of professionals, during normal deliveries or scheduled C-sections, over the distance from the 5-years moving average of number of births and 500. First of all we can notice how differences in health professional availability, observed from raw data shown in Table 4, no longer persist for scheduled C-sections after controlling for observable confounders, and for this reason we will focus on normal deliveries in the rest of the paper. Looking at Figure 1 it is possible to see how the probability of having a full team of specialists (panel a) increases sharply above the 500 births threshold, and by contrast the probability of having only gynecologist and midwife (panel b) decreases. Moreover, we observe a moderate increase in the probability of having only a midwife (panel c) and a negligible difference considering other categories (panel d).

– Insert Figures 2-3 here –

Looking at Figures 4-6 we can see the effect of having a number of births above the threshold on the probability of resuscitation, the Apgar score after 1 and 5 minutes, and the probability of Apgar scores at 1 and 5 minutes lower than a given threshold (7, 8 or 9) or equal to the perfect score (9 or 10). As we can see there is a substantial decrease in the probability of resuscitation, and an increase in the Apgar score measured after 1 minute. Moreover, there is evidence of a decrease in the probability of Apgar score at 1 minute lower than 7, of Apgar score at 5 minutes lower than 7 and 8, and an increase in the probability of perfect score after 5 minutes.

– Insert Figures 4-6 here –

4.3 Main estimates

Our main results are shown in Tables 5-6. Table 5 lists OLS estimates that do not account for endogeneity, that generally are different in sign and magnitude with respect to IV estimates.

Looking at Table 6 we can see, from our first stage regression, how the instrument is significantly different from zero and the F-statistics is above 10. A number of births above 500 increases the probability of having a full team of specialists (column 1) by almost 20 % points, which corresponds to 60% of a standard deviation. Moreover we can see how a 10 % points increase in the presence of a full team of specialists: (i) increases the probability of no need for resuscitation (column 2) by almost 0.7 % points (9% of a standard deviation), (ii) increases the Apgar score after 1 minute (column 3) by 0.076 points (8% of a standard deviation), (iii) increases the probability of having a 5 minutes Apgar score greater than 7, 8 and 9 points (columns 9-11) by 0.171 (3% of a standard deviation), 0.420 (5% of a standard deviation) and 0.857 (6% of a standard deviation) % points, respectively. No significant effects are found for other outcomes here considered.

– Insert Tables 5-6 here –

4.4 Heterogeneous effects

Now we investigate the presence of heterogeneous effects during day of week (week-ends or work days), time of day (night or day shift) or during days with a large number of deliveries (> than 75-th or 90-th percentile of the distribution of deliveries).

Tables 7-8 list estimates by day of week when OLS or IV estimators are considered. Again, we find large differences in magnitude and sign between OLS and IV. Focusing on the latter we see how the point estimate of the probability of having a full team during delivery is higher during work days (20 % points) than during week-ends (17.6 % points) even though these estimates are not significantly different, since confidence intervals widely overlap. F-statistic is close to 10 (8.9) for week-ends and well above it (31.6) during work days.

Looking at the effect on outcomes we can see how during week-ends only the probability of no need for resuscitation increases significantly. In this case we estimate that an increase of 10 % points in the probability of having a full team during delivery increases the probability of no need of resuscitation by 1.3 % points (14% of a standard deviation).

The same result holds also when we look at work days, in this case the increase is of 0.4 % points (5% of a standard deviation). In addition also the Apgar score after 1 minute increases significantly by 0.08 points (9% of a standard deviation) and we estimate a significant increase in the probability of having apgar score above 8 and 9 of 0.35 (7% of a standard deviation) and 0.9 % points (12% of a standard deviation), respectively.

– Insert Tables 7-8 here –

Looking at Tables 9 and 10 we find estimates by time of day. We considered three shifts: (i) a night shift from 12 pm to 08.59 am in the morning, (ii) a first day shift from 9 am to 5.59 pm, and (iii) a second day shift from 6 pm to 11.59 pm.

Focusing on first stages we can see how there are large differences between night and day shifts. During the night shift the full team presence increases by just 9 % points, whereas during day shifts it increases by 27 and 23 % points, respectively. F-statistics are always above 10 ranging from 10 for night shifts to 28 for the first day shift.

Looking at outcomes we find no statistically significant effects during night shifts, which is consistent with the fact that during these work shifts the probability of having a full team decreases significantly. The only statistically significant effect is found looking at the probability of no need for resuscitation during the first day shift (from 9 am to 5.59 pm), whereas we find results similar to those obtained in our baseline estimates (Table 6) for the second day shift (from 6 pm to 11.59 pm).

– Insert Tables 9-10 here –

Lastly, we estimated the effect of having a full team of specialists during days with number of births above the 75-th or 90-th percentiles of the distribution of births.

Focusing on first stage estimates we can see how, again, the probability of having a full team of specialists increases by almost 14 and 15 % points during days with number of deliveries above the 75-th and 90-th percentiles, respectively. It is interesting to notice how this figure is significantly lower than the variation of almost 20% points estimated in our baseline specification and from that estimated during day shifts (28 and 23 % points). A possible explanation could be related to the fact that if hospitals below the threshold are able to anticipate periods of high affluence they may request the presence of an additional pediatrician.

– Insert Tables 11-12 here –

4.5 Effects on maternal health

As an additional analysis we estimate the relationship between health personnel productivity and maternal health outcomes. Our dataset contains some information about maternal health and in particular we can define an indicator for the probability of having no vaginal tears of first, second and third degree, the probability of not having an episiotomy, and a measure for the absence of other complications, measured as a length of stay in hospital lower than 2 days (i.e., the standard length of stay for a normal delivery without complications).

Figure 7 provides a graphical analysis of these additional outocomes, similar to that presented in section 4.1 for newborns health. We can see that the only outcome that shows a potentially significant variation above the 500 births threshold is represented by the probability of not having 3-rd degree vaginal tears, and for this reason we will present estimates only for this outcome.

Table 13 displays second stages estimates for the whole sample and the subsamples discussed in previous sections. We estimate a significant increase in the probability of not having 3-rd degree vaginal tears in the whole sample and during work days, the first day shift (9 am to 5.59 pm) and during days with number of deliveries greater than the 90-th percentile.

– Insert Figure 7 here –

– Insert Table 13 here –

4.6 Robustness

In this section we analyse the robustness of our results proposing a series of overidentification tests, to check that no other variables potentially correlated to newborns' health status, vary discontinuously around the threshold. We also perform a robustness to check how our results change after excluding deliveries where only a midwife was present, since we saw from Figure 1 that also this category increases discontinuously after the threshold. Finally we propose a placebo test for the location of the discontinuity at various cut-off distances for our treatment and for selected outcome variables (i.e., those showing significant variations from our baseline model).

Table 14 shows a placebo analysis considering other newborns' health outcomes that should not be affected by having a number of births above the 500 births threshold. The outcomes considered are: (i) weight at birth, (ii) probability of low birthweight (weight at birth $<$ 2,500 gr) and (iii) gestational age. As we can see from Table 14 none of these vary discontinuously around the threshold, confirming that only health outcomes that should be affected by increases in productivity of health personnel are affected significantly. Table 15 shows how the presence of other teams vary around the threshold. As expected the presence of teams composed by gynecologist and midwife decreases significantly, in each sample considered, and as regards other team types we see that only during weekends the presence of teams composed by midwives increase significantly (4 % points).

Table 16 reports overidentification tests on mother's socio-demographic characteristics, and highlights how only during the second day shift (from 5 pm to 11.59 pm) there seems to be an increase in the presence of Italian mothers, a decrease in the presence of mothers with primary or secondary education and an increase of mothers with master degree. Lastly, looking at indicators of the risk associated to deliveries we do not observe a statistically significant difference in the actual level of risk of deliveries when we use other indicators, such as the probability of having an altered partogram, or altered fetal heartbeat. No significant differences are found even when we checked for differences in the probability of executing partograms or fetal heartbeat monitoring.

– Insert Table 14 here –

Since this category seems to vary discontinuously at the 500 births threshold its inclusion may bias our baseline results, so we decided to perform a robustness analysis excluding from our sample deliveries with only midwives present. Table 17 shows second stage estimates for the whole sample and for all the subsamples described in previous sections. As we can see from this Table results are very similar to those already obtained, instead, we find evidence of significant improvements also in the probability of Apgar score after 1 minute lower than 7.

Finally, if different cut offs lead to estimates of greater health personnel presence or better newborns' health improvements we would cast doubts on the effectiveness of the identification strategy, implying that other unobservable variables may be affecting our results. Figure 8 shows the placebo test for the location of the discontinuity at various cut-offs obtained varying the threshold of 500 births. We find that, consistent with our expectations, the strongest change in the key variables of our model appears exactly when the threshold is set to 500.

– Insert Figure 7 here –

5 Conclusions

In this paper we try to quantify the effect of different team compositions in the delivery room on the general health status of newborns. We find that a full team (midwife, gynecologist and pediatrician) is associated to better health outcomes. We are able to qualify this result as causal as we account for the likely reverse causality of the team composition, by instrumenting it. We exploit the legislative thresholds defining personnel requirements as instrumental variables.

Policy implications are twofold. On the one side, the association between volumes and outcomes is confirmed, as larger hospitals can ensure better health outcomes and larger productivity. However, on the other side, we believe that the actual legislation does not account for the fact that defining sharp thresholds may distort incentives to invest. The strong incentive to invest in health personnel of hospitals above thresholds may disguise the fact that hospitals right below have very similar characteristics in terms of workloads. This mechanism could favor inequalities of opportunity for women who use maternity services.

References

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

Black, S. E., Devereux, P. J. & Salvanes, K. G. (2007), 'From the Cradle to the Labor

Market? The Effect of Birth Weight on Adult Outcomes', *The Quarterly Journal of Economics* **122**(1), 409–439.

Bloom, N., Sadun, R. & Reenen, J. V. (2014), 'Does Management Matter in Healthcare?', *Harvard Business School* .

Chandra, A., Finkelstein, A., Sacarny, A. & Syverson, C. (2016), 'Productivity Dispersion in Medicine and Manufacturing', *American Economic Review: Papers Proceedings* **106**(5), 99–103.

Chandra, A. & Staiger, D. (2007), 'Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks', *Journal of Political Economy* **115**, 103–140.

Chandra, A., Staiger, D. & Skinner, J. (2010), Saving Money and Lives, in The Healthcare Imperative: Lowering Costs and Improving Outcomes, Technical report, Institute of Medicine.

URL: *http://www.ncbi.nlm.nih.gov/books/NBK53920*

Figlio, D., Guryan, J., Karbownik, K. & Roth, J. (2014), 'The Effects of Poor Neonatal Health on Children's Cognitive Development', *American Economic Review* **104**(12), 3921–3955.

Gaynor, M., Seider, H. & Vogt, W. B. (2005), 'The volume-outcome effect, scale economies, and learning-by-doing', *The American Economic Review* **95**(2), 243–247.

URL: *http://www.jstor.org/stable/4132825*

Giancotti, M., Guglielmo, A. & Mauro, M. (2017), 'Efficiency and optimal size of hospitals: Results of a systematic search', *PLoS One* **12**(3), 243–247.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–75.

La Stampa (2016), Il governo chiude i punti nascita con meno di 500 parti allanno, Technical report.

Marschak, J. & Andrews, W. (1944), 'Random Simultaneous Equations and the Theory of Production', *Econometrica* **12**(1), 143.

Ministero della salute (2010), Gli standard per la valutazione dei punti nascita, Technical report.

Oreopoulos, P., Stabile, M., Walld, R. & Roos, L. L. (2008), 'Short-, medium-, and long-term consequences of poor infant health: An analysis using siblings and twins', *Journal of Human Resources* **43**(1).

Regione Piemonte (2012), Piano Socio Sanitario 2012-2015, Technical report.

Regione Piemonte (2016), Fabbisogno Personale Ospedaliero, Technical report.

Regione Piemonte (2017), Piano Socio Sanitario 2017-2020, Technical report.

Snowden, J., Cheng, Y., Kontgis, C. & Caughey, A. (2012), 'The association between hospital obstetric volume and perinatal outcomes in California', *American Journal of Obstetrics and Gynecology* **207**(6), 478 e1–e7.

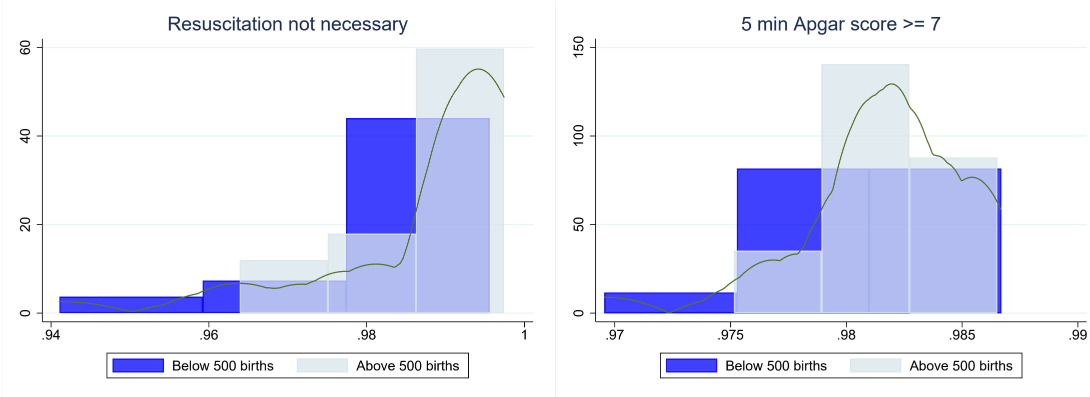

Figure 1: Productivity dispersion above and below 500 births

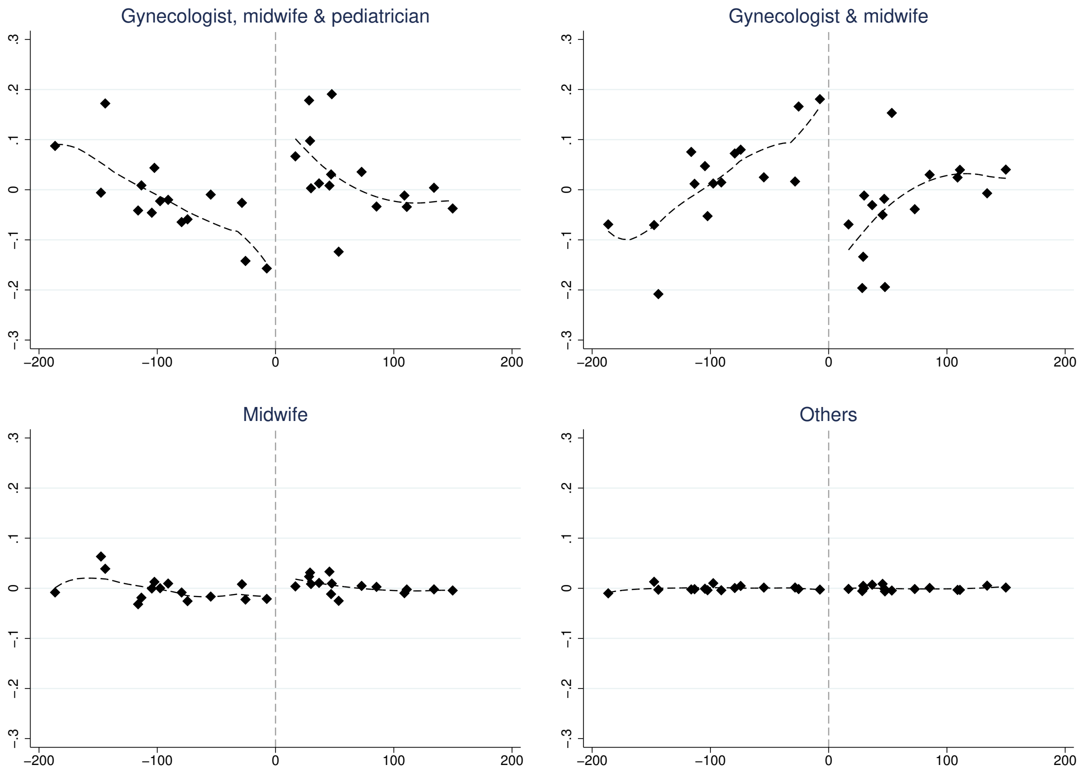

Figure 2: Residuals of regression of presence of health personnel during normal deliveries vs. maternal, delivery and hospital characteristics.

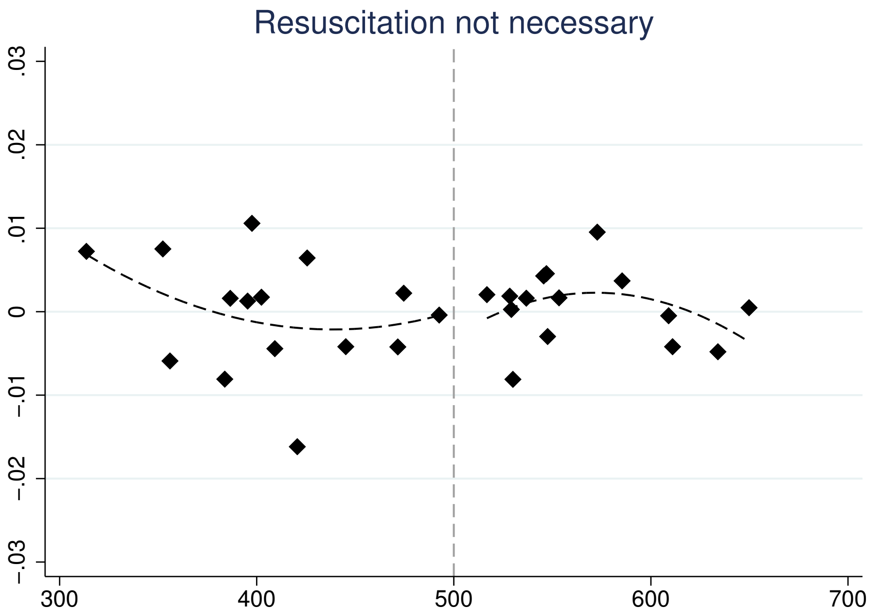

Figure 3: Residuals of regression of probability of resuscitation during normal deliveries on maternal, delivery and hospital characteristics.

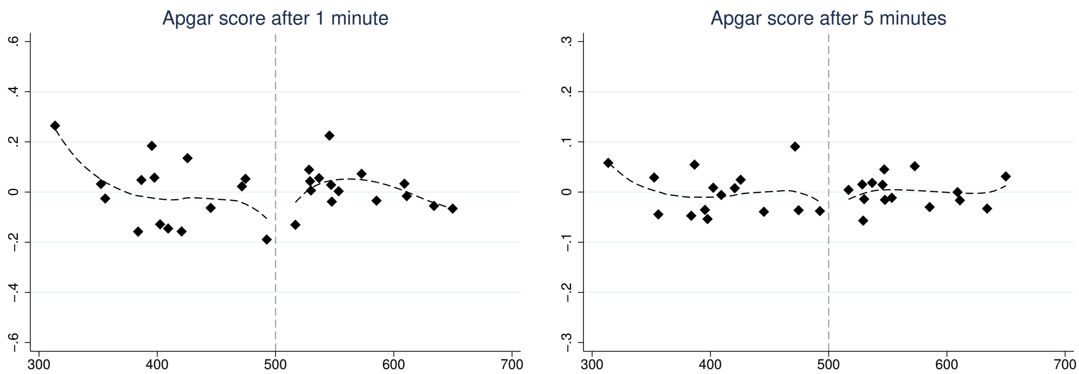

Figure 4: Residuals of regression of Apgar at 1 and 5 minutes during normal deliveries on maternal, delivery and hospital characteristics.

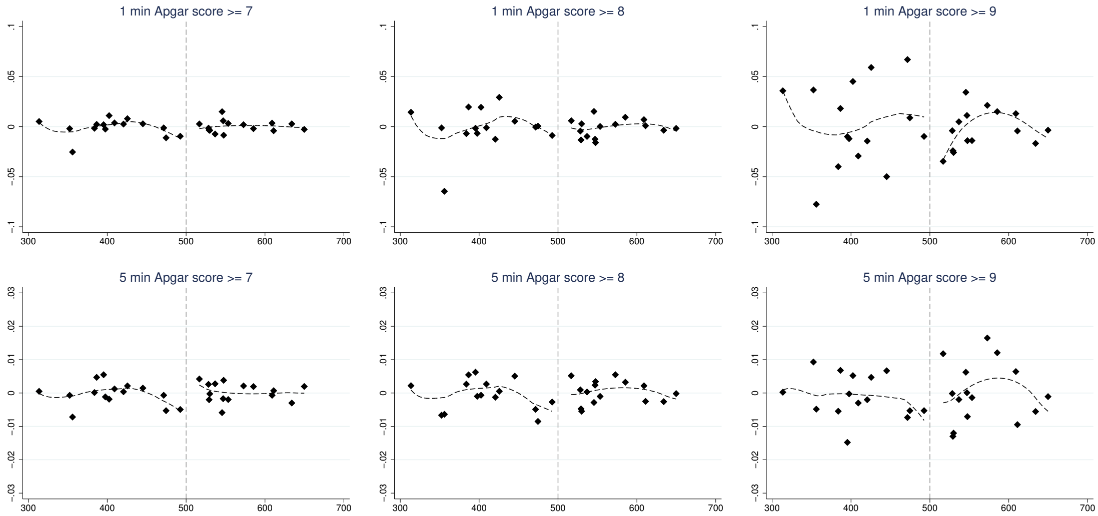

Figure 5: Residuals of regression of $Pr(apgar_i < l)$ with $l = 7, 8, 9$ and $Pr(apgar_i \geq m)$ with $m = 9, 10$ during normal deliveries on maternal, delivery and hospital characteristics.

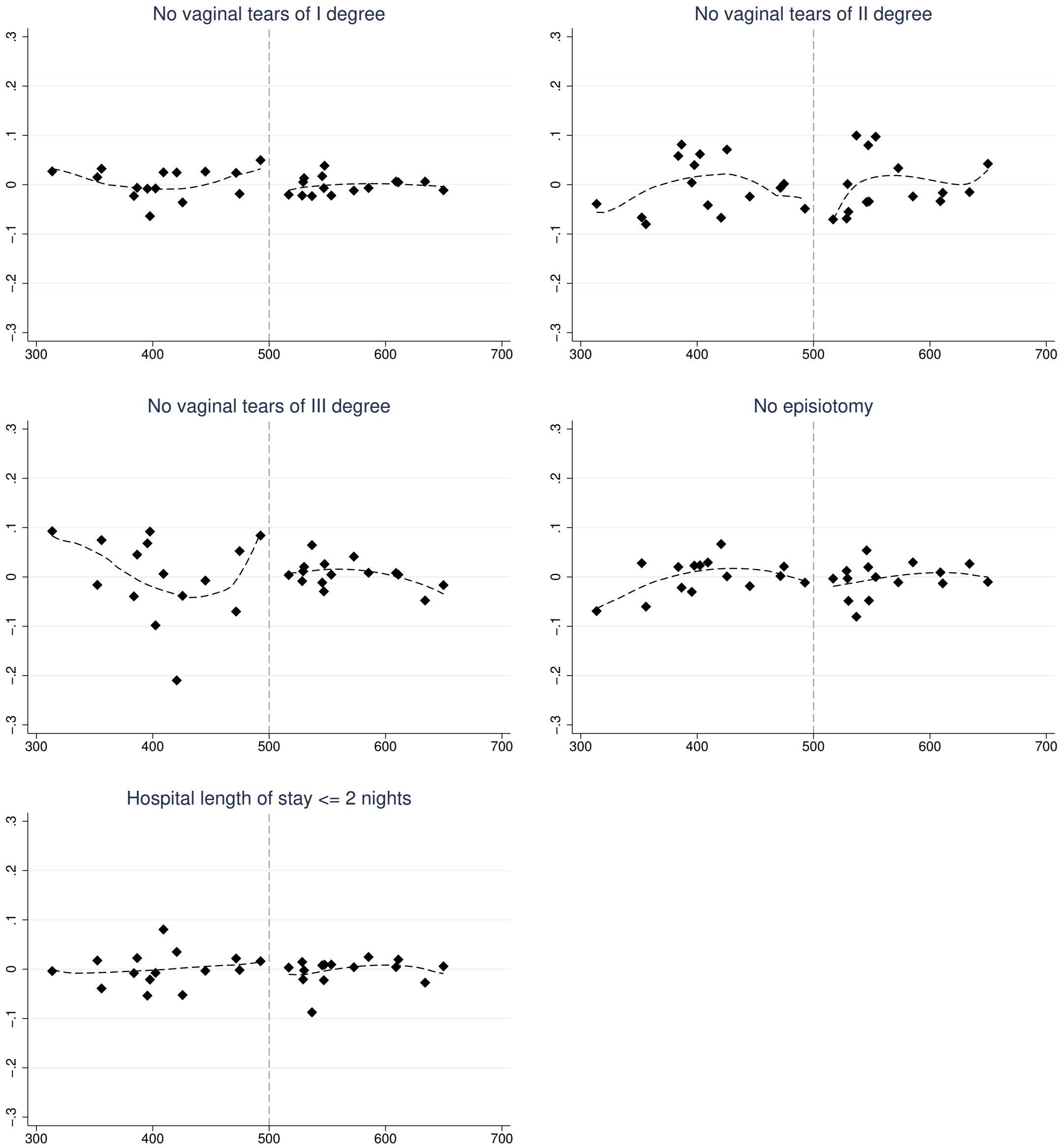

Figure 6: Residuals of regression of vaginal tears, episiotomy and length of stay in hospital during normal deliveries on maternal, delivery and hospital characteristics.

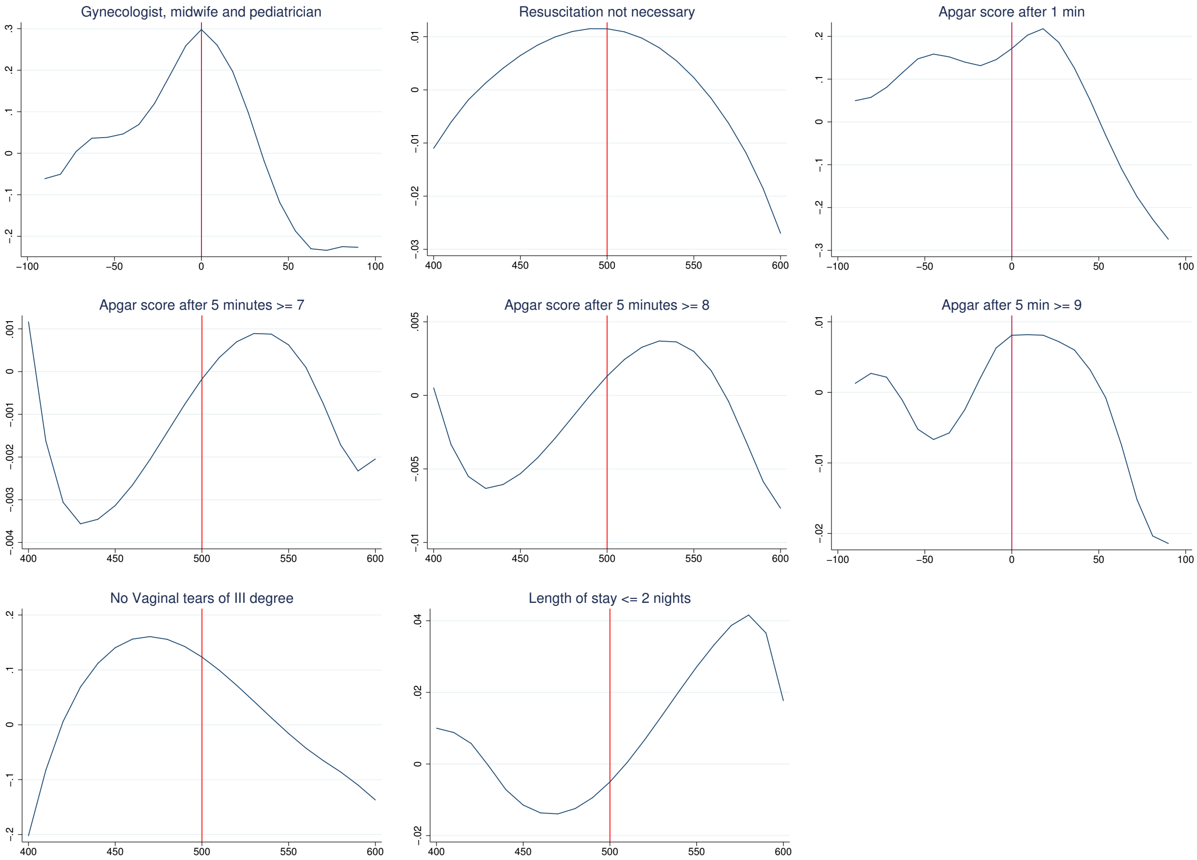

Figure 7: Placebo tests for the location of the discontinuity.

Table 1: Health personnel legislative requirements

	Staff members per work shift		
Numb. deliveries per year	**Midwife**	**Gynecologist**	**Pediatrician**
< 500	-	-	-
500 - 999	2	24/7	24/7
1,000-1,499	3	24/7	24/7
1,500-1,999	4	24/7	24/7
> 2,000	+ 1 every 750 deliveries	24/7	24/7

Source: State-Regions conference (16/12/2010).

Table 2: Number of open public competitions for recruitment of health personnel in public maternity wards, by number of births and year, in Piedmont region.

	2008-2010	**2011-2013**	**2014-2016**	**2017-2018**	**Total**
		Gynecologist			
< 500	17	1	4	23	41
500-999	19	9	8	9	37
1,000-1,499	16	5	4	9	30
1,500-1,999	13	2	4	6	21
> 2,000	7	3	2	1	11
Total	72	20	22	48	140
		Midwife			
< 500	3	0	0	3	6
500-999	19	2	7	0	21
1,000-1,499	10	1	0	1	12
1,500-1,999	21	2	4	22	45
> 2,000	2	1	1	8	11
Total	55	6	12	34	95
		Pediatrician			
< 500	24	2	11	24	50
500-999	15	26	28	22	63
1,000-1,499	23	6	8	10	39
1,500-1,999	22	5	7	8	35
> 2,000	9	6	8	3	18
Total	93	45	62	67	205

Source: own elaborations on data from the archive on public competitions in Italy and National Outcome Plan.

Table 3: Descriptive statistics

Variables	Modalities	Obs	Mean	%	Std. Dev.	Obs	Mean	%	Std. Dev.
		Whole sample				Discontinuity sample			
				Outcome measures					
Probability of no resuscitation		15,851		99.03	9.78	13,161		99.26	8.55
Apgar after 1 minute	score	15,522	8.99		100.21	12,880	9.04		99.82
	≥ 7	15,522		99.73	14.89	12,880		97.93	14.25
	≥ 8	15,522		94.81	22.18	12,880		95.05	21.68
	≥ 9	15,851		80.69	39.48	13,161		899.52	39.64
Apgar after 5 minutes	score	15,521	9.61		65.08	12,881	9.68		63.39
	≥ 7	15,521		99.54	6.8	12,881		99.57	6.58
	≥ 8	15,521		99.04	9.75	12,881		99.1	9.45
	≥ 9	15,851		97.29	16.23	13,161		97.32	16.16
				Covariates					
Mother's age		16,519	31.2		5.5	13,731	31.08		5.54
Mother's nationality	Foreign	16,543		0.27	0.45	13,751		0.29	0.45
	Italian	16,543		0.73	0.45	13,751		0.71	0.45
Mother's education	No formal education	16,460		0.01	0.08	13,671		0.01	0.08
	Compulsory education	16,460		0.02	0.12	13,671		0.02	0.13
	Primary school	16,460		0.27	0.45	13,671		0.28	0.45
	Secondary school	16,460		0.42	0.49	13,671		0.42	0.49
	Professional school	16,460		0.09	0.28	13,671		0.09	0.28
	Master degree	16,460		0.2	0.4	13,671		0.18	0.39
Mother's employment status	Unemployed or looking for a job	16,526		0.37	0.48	13,735		0.38	0.49
	Employed	16,526		0.63	0.48	13,735		0.62	0.49
Marital status	Unmarried, divorced or widow	16,492		0.34	0.47	13,704		0.34	0.47
	Married	16,492		0.66	0.47	13,704		0.66	0.47
Mother's previous deliveries	Yes	16,543		0.5	0.5	13,751		0.5	0.5
	No	16,543		0.5	0.5	13,751		0.5	0.5
Pediatrician	24/7	16,543		0.24	0.43	13,751		0.13	0.34
	Upon call	16,543		0.13	0.33	13,751		0.15	0.36
	No	16,543		0.64	0.48	13,751		0.72	0.45
Hospital level	I	16,543		1	0	13,751		1	0
	II	16,543		0	0	13,751		0	0
Gestational age	¡28	16,542		0	0.02	13,750		0	0.02
	28-31	16,542		0	0.03	13,750		0	0.04
	32-33	16,542		0	0.05	13,750		0	0.04
	34-36	16,542		0.04	0.19	13,750		0.04	0.2
	37-40	16,542		0.78	0.41	13,750		0.78	0.42
	> 40	16,542		0.18	0.38	13,750		0.18	0.38
Type of delivery	Normal	16,529		0.64	0.48	13,742		0.64	0.48
	Scheduled C-section	16,529		0.26	0.44	13,742		0.27	0.44
	Urgent C-section	16,529		0.06	0.24	13,742		0.06	0.24
	Forceps/Ventouse	16,529		0.04	0.19	13,742		0.04	0.19

Table 4: Health personnel and number of births, by type of delivery.

	350-499 births		500-650 births	
	% of personnel	**Number of deliveries**	**% of personnel**	**Number of deliveries**
		Normal deliveries		
Gynecologist, midwife and pediatrician	6.15	220	17.83	871
Gynecologist and midwife	80.18	2,868	67.11	3,279
Midwife	12.75	456	13.96	682
Other combinations	0.92	33	1.1	54
Total	100	3,577	100	4,886
		Scheduled C-sections		
Gynecologist, midwife and pediatrician	39.11	603	60.76	1,121
Gynecologist and pediatrician	57.85	892	26.88	496
Gynecologist	1.75	27	8.83	163
Other combinations	1.3	47	3.52	65
Total	100	1,542	100	1,845
		Urgent C-sections		
Gynecologist, midwife and pediatrician	46.13	185	76.08	299
Gynecologist and pediatrician	49.13	197	20.36	80
Gynecologist and midwife	1.75	7	1.53	6
Other combinations	2.99	12	2.03	8
Total	100	401	100	393
		Ventouse/Forceps		
Gynecologist, midwife and pediatrician	36.07	79	77.26	214
Gynecologist and midwife	62.1	136	22.38	62
Midwife and pediatrician	0.46	1	0.36	1
Other combinations	1.37	3	0	0
Total	100	219	100	277

Table 5: Effect of having a full team during delivery on probability of no need of resuscitation and apgar scores - OLS estimates with lha fixed effects.

	Prob of no resusc	**Apgar after 1 minute**				**Apgar after 5 minutes**			
		score	**≥ 7**	**≥ 8**	**≥ 9**	**score**	**≥ 7**	**≥ 8**	**≥ 9**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
full_team	-0.0163***	-0.5193***	-0.0587***	-0.1062***	-0.1790***	0.2340***	-0.0125***	-0.0254***	-0.0589***
	(0.005)	(0.049)	(0.009)	(0.014)	(0.017)	(0.028)	(0.004)	(0.006)	(0.010)
Mean of Y	0.55	9.14	1.51	3.78	16.02	9.72	0.31	0.63	2.04
SD of Y	7.43	0.92	12.23	19.09	36.69	0.57	5.59	7.97	14.13
Observations	8,392	8,392	8,392	8,392	8,392	8,392	8,392	8,392	8,392
Number of hospitals	10	10	10	10	10	10	10	10	10
Number of lhas	7	7	7	7	7	7	7	7	7

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 6: Effect of having a full team during delivery on probability of no need of resuscitation and apgar scores - IV estimates with lha fixed effects.

	full_team	**Prob of no resusc**	**Apgar after 1 minute**				**Apgar after 5 minutes**			
			Score	**$\geq$ 7**	**$\geq$ 8**	**$\geq$ 9**	**Score**	**$\geq$ 7**	**$\geq$ 8**	**$\geq$ 9**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Above_threshold	19.748***									
	(3.467)									
full_team		0.0668***	0.0076***	0.0330	-0.0272	0.0591	0.0013	0.0171*	0.0420*	0.0857**
		(0.026)	(0.0027)	(0.035)	(0.046)	(0.104)	(0.001)	(0.010)	(0.023)	(0.042)
Mean of Y	12.88	0.55	9.14	1.51	3.78	16.02	9.72	0.31	0.63	2.04
SD of Y	33.5	7.43	0.92	12.23	19.09	36.69	0.57	5.59	7.97	14.13
F-Stat	33.16									
Observations	8,394									
Number of hospitals	10									
Number of lhas	7									

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 7: Effect of having a full team during delivery on probability of no need of resuscitation and apgar scores by day of week - OLS estimates with lha fixed effects.

	Prob of no resusc	**Apgar after 1 minute**				**Apgar after 5 minutes**			
		score	$\geq$ **7**	$\geq$ **8**	$\geq$ **9**	**score**	$\geq$ **7**	$\geq$ **8**	$\geq$ **9**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Week-ends									
$full\_team$	-0.0165	-0.5829***	-0.0697***	-0.1093***	-0.2097***	-0.2856***	-0.0146	-0.0363**	-0.0804***
	(0.011)	(0.127)	(0.021)	(0.033)	(0.037)	(0.062)	(0.009)	(0.014)	(0.022)
Mean of Y	0.77	9.13	1.79	4.16	16.31	9.72	0.48	0.83	2.49
SD of Y	8.77	0.97	13.29	19.99	36.96	0.58	6.93	9.09	15.61
Observations	2,295	2,295	2,295	2,295	2,295	2,295	2,295	2,295	2,295
Number of hospitals	10	10	10	10	10	10	10	10	10
Number of lhas	7	7	7	7	7	7	7	7	7
Work days									
$full\_team$	-0.0164***	-0.4935***	-0.0544***	-0.1048***	-0.1677***	-0.2152***	-0.0117***	-0.0213***	-0.0513***
	(0.005)	(0.050)	(0.010)	(0.015)	(0.019)	(0.031)	(0.004)	(0.006)	(0.010)
Mean of Y	0.47	9.15	1.41	3.64	15.91	9.73	0.24	0.56	1.86
SD of Y	6.86	0.9	11.81	18.74	36.59	0.56	89.79	4.99	7.5
Observations	6,097	6,097	6,097	6,097	6,097	6,097	6,097	6,097	6,097
Number of hospitals	10	10	10	10	10	10	10	10	10
Number of lhas	7	7	7	7	7	7	7	7	7

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 8: Effect of having a full team during delivery on probability of no need of resuscitation and apgar scores by day of week - IV estimates with lha fixed effects.

	full_team	**Prob of no resusc**	**Apgar after 1 minute**				**Apgar after 5 minutes**			
			Score	**$\geq$ 7**	**$\geq$ 8**	**$\geq$ 9**	**Score**	**$\geq$ 7**	**$\geq$ 8**	**$\geq$ 9**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Week-ends										
Above_threshold	17.589**									
	(7.198)									
full_team		0.1285**	0.0055	0.0064	-0.0697	-0.0904	0.0005	0.0035	0.0546	0.0619
		(0.062)	(0.005)	(0.063)	(0.051)	(0.148)	(0.003)	(0.022)	(0.076)	(0.080)
Mean of Y	12.12	0.77	9.13	1.79	4.16	16.31	9.72	0.48	0.83	2.49
SD of Y	32.65	8.77	0.97	13.29	19.99	36.96	0.58	6.93	9.09	15.61
F-Stat	8.946									
Observations	2,296									
Number of hospitals	10									
Number of lhas	7									
Work days										
Above_threshold	20.567***									
	(3.667)									
full_team		0.0467*	0.0082***	0.0384	-0.0173	0.1084	0.0015	0.0192	0.0350**	0.0926*
		(0.025)	(0.003)	(0.027)	(0.042)	(0.113)	(0.002)	(0.017)	(0.017)	(0.049)
Mean of Y	13.17	0.47	9.15	1.41	3.64	15.91	9.73	0.24	0.56	1.86
SD of Y	33.82	6.86	0.9	11.81	18.74	36.59	0.56	89.79	4.99	7.5
F-Stat	31.59									
Observations	6,098									
Number of hospitals	10									
Number of lhas	7									

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 9: Effect of having a full team during delivery on probability of no need of resuscitation and apgar scores by time of day - OLS estimates with lha fixed effects.

	Prob of	**Apgar after 1 minute**				**Apgar after 5 minutes**			
	no resusc	**score**	$\geq$ **7**	$\geq$ **8**	$\geq$ **9**	**score**	$\geq$ **7**	$\geq$ **8**	$\geq$ **9**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Night shift (from 12 pm to 8.59 am)									
full_team	-0.0160	-0.4877***	-0.0575***	-0.0946***	-0.1933***	-0.2392***	-0.0191**	-0.0254**	-0.0758***
	(0.010)	(0.110)	(0.019)	(0.024)	(0.035)	(0.068)	(0.010)	(0.010)	(0.020)
Mean of Y	0.49	9.18	1.35	3.44	15.38	9.73	0.31	0.51	2.05
SD of Y	7.04	0.88	11.58	18.24	36.09	0.58	5.61	7.09	14.18
Observations	3,188	3,188	3,188	3,188	3,188	3,188	3,188	3,188	3,188
Number of hospitals	10	10	10	10	10	10	10	10	10
Number of lhas	7	7	7	7	7	7	7	7	7
Day shift I (from 9 am to 5.59 pm)									
full_team	-0.0141**	-0.4876***	-0.0527***	-0.1093***	-0.1702***	-0.2222***	-0.0079	-0.0218***	-0.0446***
	(0.006)	(0.070)	(0.012)	(0.018)	(0.026)	(0.039)	(0.005)	(0.008)	(0.011)
Mean of Y	0.62	9.12	1.52	3.88	16.11	9.72	0.32	0.74	1.76
SD of Y	7.87	0.94	12.23	19.33	36.76	0.57	5.63	8.58	13.17
Observations	2,866	2,866	2,866	2,866	2,866	2,866	2,866	2,866	2,866
Number of hospitals	10	10	10	10	10	10	10	10	10
Number of lhas	7	7	7	7	7	7	7	7	7
Day shift II (from 6 pm to 11.59 pm)									
full_team	-0.0180*	-0.5879***	-0.0727***	-0.1185***	-0.1850***	-0.2559***	-0.0143**	-0.0314***	-0.0678***
	(0.010)	(0.091)	(0.016)	(0.023)	(0.025)	(0.047)	(0.007)	(0.010)	(0.017)
Mean of Y	0.55	9.13	17.42	4.13	16.81	9.73	0.31	0.69	2.35
SD of Y	7.42	0.94	13.09	19.92	37.41	0.56	5.51	8.32	15.16
Observations	2,338	2,338	2,338	2,338	2,338	2,338	2,338	2,338	2,338
Number of hospitals	10	10	10	10	10	10	10	10	10
Number of lhas	7	7	7	7	7	7	7	7	7

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$*

Table 10: Effect of having a full team during delivery on probability of no need of resuscitation and apgar scores by time of day - IV estimates with lha fixed effects.

	full_team	**Prob of no resusc**	**Apgar after 1 minute**				**Apgar after 5 minutes**			
			Score	**$\geq$ 7**	**$\geq$ 8**	**$\geq$ 9**	**Score**	**$\geq$ 7**	**$\geq$ 8**	**$\geq$ 9**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
				Night shift (from 12 pm to 8.59 am)						
Above_threshold	9.614***									
	(2.876)									
full_team		0.1552	0.00154	0.0293	-0.0237	0.2311	0.0019	-0.0109	-0.0175	0.0846
		(0.113)	(0.010)	(0.112)	(0.125)	(0.447)	(0.004)	(0.046)	(0.074)	(0.096)
Mean of Y	7.89	0.49	9.18	1.35	3.44	15.38	9.73	0.31	0.51	2.05
SD of Y	26.97	7.04	0.88	11.58	18.24	36.09	0.58	5.61	7.09	14.18
F-Stat	10.57									
Observations	3,190									
Number of hospitals	10									
Number of lhas	7									
				Day shift I (from 9 am to 5.59 pm)						
Above_threshold	27.524***									
	(5.236)									
full_team		0.0360*	0.0028	-0.0224	-0.0735	0.0085	0.0006	0.0151	0.0376	0.0359
		(0.021)	(0.003)	(0.033)	(0.056)	(0.099)	(0.002)	(0.023)	(0.025)	(0.046)
Mean of Y	16.92	0.62	9.12	1.52	3.88	16.11	9.72	0.32	0.74	1.76
SD of Y	37.5	7.87	0.94	12.23	19.33	36.76	0.57	5.63	8.58	13.17
F-Stat	28.18									
Observations	2,866									
Number of hospitals	10									
Number of lhas	7									
				Day shift II (from 6 pm to 11.59 pm)						
Above_threshold	23.606***									
	(4.675)									
full_team		0.0591**	0.0011**	0.1130	0.0209	0.0563	0.0037	0.0373*	0.0896*	0.1581
		(0.028)	(0.005)	(0.077)	(0.095)	(0.154)	(0.003)	(0.022)	(0.051)	(0.098)
Mean of Y	14.75	0.55	9.13	17.42	4.13	16.81	9.73	0.31	0.69	2.35
SD of Y	35.47	7.42	0.94	13.09	19.92	37.41	0.56	5.51	8.32	15.16
F-Stat	22.30									
Observations	2,338									
Number of hospitals	10									
Number of lhas	7									

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 11: Effect of having a full team during delivery on probability of no need of resuscitation and apgar scores during crowded days - OLS estimates with lha fixed effects.

	Prob of no resusc	Apgar after 1 minute score	Apgar after 1 minute ≥ 7	Apgar after 1 minute ≥ 8	Apgar after 1 minute ≥ 9	Apgar after 5 minutes score	Apgar after 5 minutes ≥ 7	Apgar after 5 minutes ≥ 8	Apgar after 5 minutes ≥ 9
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
	Daily number of births > 75-th percentile								
$full\_team$	-0.0171**	-0.7257***	-0.1002***	-0.1661***	-0.2350***	0.3157***	-0.0101	-0.0334***	-0.0937***
	(0.008)	(0.093)	(0.017)	(0.022)	(0.028)	(0.055)	(0.006)	(0.012)	(0.018)
Mean of Y	0.43	9.13	14.86	3.96	16.25	9.69	0.23	0.61	2.21
SD of Y	49.34	0.92	12.1	19.51	36.9	0.58	4.82	7.8	14.72
Observations	3,498	3,498	3,498	3,498	3,498	3,498	3,498	3,498	3,498
Number of hospitals	10	10	10	10	10	10	10	10	10
Number of lhas	7	7	7	7	7	7	7	7	7
	Daily number of births > 90-th percentile								
$full\_team$	-0.0179*	-0.7426***	-0.0852***	-0.1782***	-0.2261***	0.3114***	-0.0081	-0.0345**	-0.0939***
	(0.010)	(0.115)	(0.019)	(0.026)	(0.033)	(0.064)	(0.007)	(0.014)	(0.022)
Mean of Y	0.43	9.24	1.24	3.39	13.46	9.76	0.16	0.56	17.17
SD of Y	6.54	0.9	11.06	18.12	34.14	0.53	3.99	7.46	12.99
Observations	2,552	2,552	2,552	2,552	2,552	2,552	2,552	2,552	2,552
Number of hospitals	10	10	10	10	10	10	10	10	10
Number of lhas	7	7	7	7	7	7	7	7	7

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 12: Effect of having a full team during delivery on probability of no need of resuscitation and apgar scores during crowded days - IV estimates with lha fixed effects.

	full_team	**Prob of no resusc**	**Apgar after 1 minute**				**Apgar after 5 minutes**			
			Score	**$\geq$ 7**	**$\geq$ 8**	**$\geq$ 9**	**Score**	**$\geq$ 7**	**$\geq$ 8**	**$\geq$ 9**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
				Daily number of births > 75-th percentile						
Above_threshold	13.907***									
	(4.367)									
full_team		0.0642***	0.0086	0.0119	-0.0955	0.1204	-0.0001	0.0035	0.0065	0.0156
		(0.024)	(0.006)	(0.089)	(0.107)	(0.182)	(0.004)	(0.037)	(0.056)	(0.085)
Mean of Y	9.47	0.43	9.13	14.86	3.96	16.25	9.69	0.23	0.61	2.21
SD of Y	29.29	49.34	0.92	12.1	19.51	36.9	0.58	4.82	7.8	14.72
F-Stat	10.40									
Observations	3,499									
Number of hospitals	10									
Number of lhas	7									
				Daily number of births > 90-th percentile						
Above_threshold	14.901***									
	(5.234)									
full_team		0.0883*	0.013*	-0.0112	-0.0461	0.2379	-0.0032	0.0267	0.0290	0.0532
		(0.046)	(0.008)	(0.097)	(0.085)	(0.243)	(0.004)	(0.038)	(0.077)	(0.090)
Mean of Y	9.11	0.43	9.24	1.24	3.39	13.46	9.76	0.16	0.56	17.17
SD of Y	28.64	6.54	0.9	11.06	18.12	34.14	0.53	3.99	7.46	12.99
F-Stat	9.11									
Observations	3,499									
Number of hospitals	10									
Number of lhas	7									

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 13: Effect of having a full team during delivery on probability of 3-rd degree vaginal tears and of having length of stay in hospital longer than 2 days - IV estimates with lha fixed effects.

	All	**Week-ends**	**Work days**	**Night shift**	**Day shift I**	**Day shift II**	**Daily number of births**	
				(12 pm-8.59 am)	(9 am-4.59 pm)	(5 pm-11.59 pm)	> 75-th perc	> 90-th perc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				No vaginal tears of 3-rd degree				
full_team	0.1976*	0.2107	0.1985*	0.4863*	0.1550*	0.1039	0.2877	0.5231*
	(0.115)	(0.287)	(0.115)	(0.291)	(0.093)	(0.185)	(0.252)	(0.281)
Mean of Y	15.53	15.65	15.48	14.94	15.71	16.11	16.67	16.85
SD of Y	36.22	36.35	36.18	35.66	36.4	36.77	37.28	37.44
				Mother's length of stay in hospital $\leq$ 2				
full_team	-0.0544*	-0.2718*	0.0160	-0.0355	-0.0157	-0.1039	-0.0308	-0.0448
	(0.032)	(0.158)	(0.234)	(0.321)	(0.080)	(0.117)	(0.126)	(0.275)
Mean of Y	56.78	55.22	57.37	57.20	57.57	55.22	56.71	56.69
SD of Y	49.54	49.73	49.45	49.48	49.43	49.73	49.55	49.56

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 14: Robustness I: Placebo outcomes.

	All	Week-ends	Work days	Night shift (12 pm-8.59 am)	Day shift I (9 am-4.59 pm)	Day shift II (5 pm-11.59 pm)	Daily number of births > 75-th perc	90-th perc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Weight at birth	-22.2199	-43.3911	-15.1639	-143.2359	40.9370	74.7228	-111.5412	-17.6010
	(115.448)	(223.583)	(135.343)	(115.128)	(125.992)	(125.876)	(136.085)	(165.327)
Weigth at birth < 2,500 gr	0.0112	0.0321	0.0039	0.0485	-0.0252	0.0008	0.0394	0.0024
	(0.034)	(0.065)	(0.040)	(0.034)	(0.036)	(0.038)	(0.039)	(0.047)
Gestational age	0.0693	0.0737	0.0623	0.0572	0.0434	0.1228	0.1674**	0.1049
	(0.047)	(0.087)	(0.053)	(0.080)	(0.078)	(0.095)	(0.076)	(0.093)

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 15: Robustness II: Other teams.

	All	Week-ends	Work days	Night shift (12 pm-8.59 am)	Day shift I (9 am-4.59 pm)	Day shift II (5 pm-11.59 pm)	Daily number of births > 75-th perc	90-th perc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Other teams			
Gynecologist and midwife	-0.2145***	-0.2161***	-0.2151***	-0.1109***	-0.2870***	-0.2544***	-0.1386***	-0.1560***
	(0.031)	(0.052)	(0.037)	(0.031)	(0.049)	(0.044)	(0.039)	(0.049)
Midwife	0.0191	0.0431***	0.0113	0.0092	0.0115	0.0345	0.0011	0.0146
	(0.015)	(0.015)	(0.018)	(0.023)	(0.021)	(0.026)	(0.022)	(0.027)
Other	-0.0007	-0.0017	-0.0005	0.0022	0.0077	-0.0140*	-0.0037	-0.0127
	(0.003)	(0.007)	(0.004)	(0.006)	(0.006)	(0.008)	(0.006)	(0.009)

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 16: Robustness III: Overidentification tests.

	All	Week-ends	Work days	Night shift (12 pm-8.59 am)	Day shift I (9 am-4.59 pm)	Day shift II (5 pm-11.59 pm)	Daily number of births > 75-th perc	90-th perc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Mother's characteristics			
Age	0.2703	0.3103	0.1979	0.5512	0.0647	0.1447	0.2188	0.5032
	(0.214)	(0.415)	(0.237)	(0.382)	(0.379)	(0.474)	(0.361)	(0.489)
Italian	0.0300	0.0485	0.0224	0.0221	-0.0180	0.0993**	-0.0058	-0.0502
	(0.022)	(0.043)	(0.025)	(0.032)	(0.042)	(0.045)	(0.033)	(0.045)
Primary or secondary education	0.0019	0.0054	0.0025	0.0340	0.0383	-0.0902***	0.0202	0.0568
	(0.015)	(0.030)	(0.018)	(0.024)	(0.030)	(0.032)	(0.029)	(0.041)
Master degree	0.0045	0.0028	0.0032	-0.0237	-0.0429	0.1034***	-0.0078	-0.0479
	(0.017)	(0.031)	(0.020)	(0.026)	(0.030)	(0.032)	(0.028)	(0.039)
Pluriaprous	0.0220	0.0164	0.0245	0.0069	0.0216	0.0363	0.0216	0.0237
	(0.021)	(0.042)	(0.024)	(0.036)	(0.034)	(0.046)	(0.034)	(0.047)
					Deliveries at risk			
Induced labor	0.0596	0.7006***	-0.2292	-0.7918	0.4508	0.3436	0.6336	2.0924
	(0.474)	(0.192)	(0.560)	(0.765)	(0.548)	(0.921)	(1.233)	(2.664)
Antibiotics during labor	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Altered partogram	0.0029	-0.0018	0.0046	-0.0007	0.0062	0.0031	0.0046	0.0289***
	(0.004)	(0.006)	(0.004)	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)
Partogram not executed	-0.0126	-0.0254	-0.0075	-0.0227	-0.0103	-0.0007	-0.0262	-0.0154
	(0.014)	(0.020)	(0.018)	(0.025)	(0.024)	(0.031)	(0.023)	(0.031)
Altered fetal heartbeat	-0.0016	-0.0020	-0.0013	0.0076	0.0003	-0.0160*	-0.0064	-0.0156
	(0.004)	(0.007)	(0.005)	(0.006)	(0.005)	(0.009)	(0.006)	(0.010)
Fetal heartbeat not executed	-0.0042	-0.0264	0.0043	0.0019	-0.0010	-0.0163	-0.0021	-0.0026
	(0.013)	(0.030)	(0.013)	(0.019)	(0.022)	(0.021)	(0.021)	(0.019)
Observations	8,460	2,325	6,135	3,218	2,890	2,352	3,514	2,563

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 17: Effect of having a full team during delivery on newborns' and maternal health status - IV estimates with lha fixed effects excluding deliveries with midwives only.

	All	Week-ends	Work days	Night shift (12 pm-8.59 am)	Day shift I (9 am-4.59 pm)	Day shift II (5 pm-11.59 pm)	Daily number of births > 75-th perc	90-th perc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No need of resuscitation	0.0695**	0.1234**	0.0498*	0.1505	0.0387*	0.0683**	0.0617***	0.0740*
	(0.027)	(0.052)	(0.028)	(0.120)	(0.022)	(0.027)	(0.023)	(0.041)
Apgar score after 1 min	0.7259**	0.4200	0.8164***	1.2387	0.2342	1.2538**	0.8859*	1.1586*
	(0.282)	(0.444)	(0.275)	(1.018)	(0.301)	(0.543)	(0.515)	(0.640)
Apgar score after 1 min ≥ 7	0.0383	0.0002	0.0465*	0.0240	-0.0252	0.1391*	0.0305	-0.0197
	(0.036)	(0.058)	(0.028)	(0.123)	(0.032)	(0.083)	(0.091)	(0.092)
Apgar score after 1 min ≥ 8	-0.0225	-0.0679	-0.0135	-0.0293	-0.0831	0.0584	-0.0614	-0.0229
	(0.048)	(0.056)	(0.043)	(0.144)	(0.058)	(0.105)	(0.114)	(0.078)
Apgar score after 1 min ≥ 9	0.0732	-0.0959	0.1277	0.1698	0.0272	0.1190	0.1330	0.2068
	(0.104)	(0.136)	(0.114)	(0.454)	(0.085)	(0.148)	(0.184)	(0.231)
Apgar score after 5 min	0.1731	0.0919	0.1918	0.1712	-0.0159	0.4480	0.1189	0.2617
	(0.141)	(0.253)	(0.155)	(0.472)	(0.165)	(0.290)	(0.336)	(0.330)
Apgar score after 5 min ≥ 7	0.0222*	0.0217	0.0201	-0.0007	0.0174	0.0460*	0.0156	0.0212
	(0.013)	(0.032)	(0.018)	(0.044)	(0.025)	(0.027)	(0.036)	(0.035)
Apgar score after 5 min ≥ 8	0.0490*	0.0730	0.0376**	-0.0061	0.0406	0.0978*	0.0163	0.0131
	(0.026)	(0.085)	(0.019)	(0.077)	(0.028)	(0.054)	(0.054)	(0.075)
Apgar score after 5 min ≥ 9	0.0995**	0.0907	0.1005*	0.0908	0.0444	0.1823*	0.0501	0.0662
	(0.046)	(0.083)	(0.054)	(0.101)	(0.048)	(0.104)	(0.101)	(0.093)
Length of stay in hospital ≤ 2	-0.0283	-0.2264	0.0459	0.0271	0.0035	-0.0714	0.0140	-0.0426
	(0.036)	(0.165)	(0.354)	(0.317)	(0.094)	(0.128)	(0.145)	(0.238)
No vaginal tears - 3 rd degree	0.1513	0.1634	0.1604	0.4078	0.1349	0.0800	0.2666	0.4366*
	(0.126)	(0.303)	(0.133)	(0.304)	(0.089)	(0.183)	(0.232)	(0.247)
Observations	7,325	2,054	5,271	2,763	2,488	2,074	3,116	2,294

*Notes: Standard errors clustered at the lha, day of week, month and year level. Significant levels: *** p<0.01, ** p<0.05, * p<0.1*