Labour Economics 96 (2025) 102760

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Understanding labour productivity in maternity wards

Marina Di Giacomo [a,*], Massimiliano Piacenza [b], Luca Salmasi [c], Gilberto Turati [c]

[a] University of Torino, Department of Economics, Social Sciences, Applied Mathematics and Statistics (ESOMAS), Torino, Italy
[b] University of Piemonte Orientale, Department of Law and Political, Economic and Social Sciences (DIGSPES), Alessandria, Italy
[c] Università Cattolica del Sacro Cuore, Department of Economics and Finance, Rome, Italy

ARTICLE INFO

JEL classification:
J24
D24
I18
H75

Keywords:
Medical staff
Maternity wards
Productivity
Instrumental variables

ABSTRACT

This paper provides a causal estimate of labour productivity in maternity wards. We consider an Italian law that defines the staffing requirements of hospital maternity units according to the annual number of births. We exploit the discontinuities in the availability of medical staff caused by thresholds in the law to define both instrumental variables and a regression discontinuity framework that allows us to estimate the causal effect of different teams of professionals on the mode of delivery and on the health status of newborns and mothers at delivery. The analysis is based on detailed patient-level data on births in an Italian region. We find that maternity units with annual births above the thresholds are more likely to have a 'full team' of professionals at delivery. We find that having a full team has no effect on the mode of delivery (caesarean section vs vaginal birth). However, the presence of a full team has a significant impact on health outcomes. We find an improvement in both neonatal and maternal outcomes associated with a more intensive use of medical interventions, suggesting that larger hospitals are better than smaller units at managing deliveries with appropriate treatments to avoid complications. In addition, we do not find substantial heterogeneous effects across days of the week, time of day, or nationality of mothers.

1. Introduction

Healthcare expenditure has been rapidly growing in the past decades and available estimates suggest it will increase further in the future almost everywhere in the world. However, there are different views on why health expenditure is growing (OECD, 2015). Demand-side explanations are mostly based on aging populations and the "time-to-death" hypothesis (e.g., Zweifel et al., 1999; Gruber and Wise, 2002). Supply-side explanations are divided between classical Baumol's disease and modern views, which suggest that health care is no different from other industries in terms of cross-sectional heterogeneity in the productivity characterizing different providers (e.g., Chandra et al., 2016). According to this explanation, health spending is plagued by large inefficiencies and waste (e.g., Chandra and Staiger, 2020; Shrank et al., 2019). Overtreatment, adoption of ineffective, costly technologies, and inappropriate treatments are all examples where there is large room for improvement. Reducing inefficiencies and increasing productivity are ways to address the increasing demand for healthcare services in the face of tightening budgets (Baicker et al., 2012; OECD, 2024). Given its importance, the issue of measuring and explaining productivity in healthcare has recently received increasing attention from scholars (e.g., Chandra and Staiger, 2007; Nicholson and Propper, 2011; Bloom et al., 2015; Skinner and Staiger, 2015; Chandra et al., 2016; Lee et al., 2019).

In this paper, we study labor productivity in hospital maternity wards in the Italian NHS, a universal tax-funded healthcare system managed at the regional level. We aim to estimate the causal effect of medical staff on the health status of newborns and their mothers. The identification strategy exploits the exogenous variation in staffing and equipment requirements imposed on maternity wards by Italian legislation according to a historic number of deliveries. The statutory thresholds defined by the law create discontinuities in the availability of professionals attending deliveries. We use these exogenous jumps to identify instrumental variables for staffing. We then use these variables in a two-stage least squares instrumental variables (2SLS-IV) setting and a regression discontinuity analysis (RDD) to estimate the causal effect of medical staff on delivery mode and a large set of neonatal and maternal health outcomes. This approach avoids the simultaneity bias arising from clinical staffing choices made in response to patient needs.

In analyzing maternity wards, we focus on *intrapartum care*, i.e., care mothers and newborns receive during labor and immediately after birth.

* Corresponding author.
E-mail addresses: marina.digiacomo@unito.it (M. Di Giacomo), massimiliano.piacenza@uniupo.it (M. Piacenza), luca.salmasi@unicatt.it (L. Salmasi), gilberto.turati@unicatt.it (G. Turati).

https://doi.org/10.1016/j.labeco.2025.102760
Received 27 February 2024; Received in revised form 30 June 2025; Accepted 7 July 2025
Available online 11 July 2025

The focus on short-term health outcomes is important for at least two reasons. First, decisions taken in the first minutes after birth, known as the "golden minute" (or the "golden hour"), can have a significant impact on long-term health outcomes both for newborns and mothers (e.g., Vento et al., 2009). This provides sharper estimates of the causal relationship between medical treatment and long-term outcomes than using measures of health in later life, which may reflect the cumulative effects of multiple treatments that have occurred in the past (e.g., Carrillo and Feres, 2019). Understanding the causal impact of medical staff on neonatal health is crucial since neonatal health is a strong predictor of later health outcomes (e.g., Almond et al., 2005), cognitive ability (e.g., Figlio et al., 2014; Bharadwaj et al., 2013), and labor market outcomes (e.g., Black et al., 2007; Almond 2006). Similarly, intrapartum maternal health can have long-lasting effects, including reduced/zero complications in the postpartum period, better mental health, and higher fertility rates (WHO, 2018; Leinweber et al., 2023). Second, labor and delivery are carried out in a unique environment characterized by the joint presence of many professional workers, including midwives, gynecologists, and pediatricians. For a single patient, the composition of a labor and delivery team can vary widely, even during a single episode, depending on individual patient needs and available clinical resources. However, according to clinical guidelines, a large team of different professionals is crucial for obtaining a good outcome (e.g., ACOG, 2017; Batey et al., 2022).

Our empirical analysis is based on a detailed patient-level dataset covering the universe of births (both vaginal and caesarean) that occurred within the administrative boundaries of Piedmont, a north-western region of Italy, between 2011 and 2013. We consider a productivity regression model in which clinical staff is the main input for the mode of delivery (the intermediate output) and several health outcomes for both newborns and mothers during and immediately after childbirth. Clinical staff availability is measured by a binary indicator that equals one if a "full team" of professionals (including at least a midwife, a gynecologist, and a pediatrician) attended the birth, capturing both the size and the composition of the team in terms of skill diversity. To properly identify the effect of full team availability and composition on health outcomes, in both the 2SLS-IV and RDD models we control for a large set of observable characteristics that account for differences in needs during delivery. We also include local health authority and time (year and month) fixed effects, which can allow for systematic differences in the quality of health services across local authorities and over time. Given the three years covered by our data and the focus on maternity units, we expect technological innovation to play a minor role in our setting.

To account for the potential endogeneity of the full team, we exploit the discontinuities in the staff and equipment requirements mandated by a set of regulations derived from the Decrees of the Ministry of Health D. M. 24/04/2000 and D.M. 14/02/2001 defining the "Project targeting mothers and infants". The law classifies maternity units into three groups based on the average number of deliveries in the previous years: (i) units with <500 deliveries per year; (ii) units with between 500 and 999 deliveries per year; and (iii) maternity units with 1000 or more deliveries per year. Based on the group to which the maternity unit belongs, the law specifies the medical staff units per shift that need to be available (in terms of midwives, gynecologists and pediatricians/neonatologists), equipment and supplies (number of obstetric and inpatient beds, number of labour and delivery rooms). We use dummy variables identifying each group as instrumental variables for the "full team" variable, but we also exploit the discontinuities in an RDD framework to complement the IV analysis.

Our identification strategy is based on the assumption that the volume group to which the maternity unit belongs affects health outcomes only through the team of professionals attending the birth. Given the critical role of the professional team in 'producing' birth health outcomes, we believe that this assumption is likely to be correct. We present evidence to support our findings through overidentification tests and the results of an RDD that compares very similar maternity units in terms of volumes treated, that due to their location relative to the critical thresholds, are subject to different regulatory requirements.

First-stage estimates support the relevance of instruments based on the exogenous cut-offs defined by the law. In our RDD samples, which include all deliveries, we find that maternity units with annual births above the thresholds are more likely to have a full team at delivery than those below the thresholds: the probability of having a full team in the delivery room increases by about 23 percentage points for maternity units above the 500-births threshold and by about 18 percentage points for units above the 1000-births threshold compared to units just below the thresholds.

As a preliminary step, we examine the effect of a full team on the mode of delivery (our intermediate output). We find no role for full team in the likelihood of having a caesarean section (C-section) rather than a vaginal delivery. We then consider the role of full team on health outcomes at birth. Our estimates suggest that a full team of professionals (compared to a smaller and less diverse team in terms of specialization) significantly improves neonatal and maternal health outcomes, especially for vaginal deliveries. Concerning neonatal outcomes, according to our 2SLS results in the sample of all births, a full team increases the probability of no need for resuscitation by 2.3 percentage points (about one-fifth of a standard deviation). Interestingly, Apgar scores are significantly lower with a full team: −0.36 points and −0.12 points for one- and five-minute Apgar scores, respectively. A full team also decreases the probability of observing a perfect Apgar score (greater than nine) after birth. However, considering the difference between the Apgar score measured at five minutes and the Apgar score at one minute from birth, we find that the presence of a full team significantly improves the five-minute score relative to the one-minute score. The five-minute score is significantly larger than the one-minute score by 0.23 points (one-fifth of a standard deviation). We interpret this result as the effect of the presence of a full team compared to a smaller team on the health status of a newborn in the few minutes that follow childbirth. It is especially likely that the presence of a pediatrician who assists with the birth may positively affect the health status because of her expertise. This indicates the full team's active role in improving neonatal health in the first few minutes after birth. For maternal health outcomes, we find that the likelihood of having no major obstetric lacerations increases by about nine percentage points (about one-quarter of a standard deviation) with a full team. At the same time, the likelihood of having an episiotomy, a small surgical cut to facilitate childbirth and prevent complications or vaginal tears, increases by 17 percentage points with a full team, suggesting that a full team performs this treatment when it is needed to prevent lacerations.

We extend our analysis in two ways. First, we replicate our estimates by splitting the sample into many categories: vaginal deliveries, assisted vaginal deliveries (requiring the use of special equipment, such as forceps and vacuum devices), uncomplicated vaginal deliveries, emergency C-sections, and planned C-sections. We find that our baseline results are driven only by vaginal deliveries, in particular uncomplicated vaginal deliveries, while there is no effect on the two C-section subsamples. We interpret this finding as evidence of different production processes for the two procedures: while a vaginal delivery is a medical procedure, a C-section is a specialized surgical procedure requiring a specific team and an operating room.

Focusing on uncomplicated vaginal births, we also discuss the RDD estimates, which allow us to compare maternity units that are very close in terms of workload but differ according to their position above or below the two thresholds set by the law. Our main findings are confirmed: vaginal births in units above the cut-off are associated with better neonatal and maternal health outcomes, and the results are consistent for the overlapping sample and the two samples around the 500 and 1000 births thresholds. The coefficients are generally larger for larger hospitals, confirming the association between volume and outcomes, with larger hospitals ensuring greater productivity as measured by better health outcomes. Finally, we examine heterogeneous effects across days of the week, time of day, and mother's nationality and find no significant heterogeneous effects.

Our study contributes to two different strands of literature. First, it is related to the growing broad literature studying labor productivity differences in the health sector (e.g., Nicholson and Propper, 2011; Lee et al., 2019, for recent surveys). The relationship between medical staff inputs and patient outcomes has been studied from different perspectives. First, many studies focus on the direct impact of staff availability on patient outcomes (e.g., Doyle et al., 2010; Gruber and Kleiner, 2012; Rogowski et al., 2013; Lin, 2014; Matsudaira, 2014; Friedrich and Hackmann, 2017; Carrillo and Feres, 2019; Einav et al., 2022; Raja, 2023). One typical – and unsurprising – result is that understaffing (especially, nurse understaffing) increases the risk of adverse health outcomes. There is also some evidence that nurses and physicians may be substitutes (Carrillo and Feres, 2019), and mandated minimum nurse-to-patient ratios have mixed effects on health outcomes (Aiken et al., 2011; Cook et al., 2012; Raja, 2023). The second perspective considers the role of health professionals' skill mix, experience, and deployment (e.g., Almond et al., 2010; Bartel et al., 2014; Chan, 2016, 2018; Daysal et al., 2019; Silver, 2020). The evidence points to large differences across professionals. Some of these works relate to studies of peer effects and find that relationships with coworkers can either improve communication and coordination, generating positive health outcomes (e.g., Chan, 2016) or trigger perverse effects negatively influencing health outcomes (e.g., Bartel et al., 2014; Chan, 2018). A third perspective exploits potential congestion episodes and hospital crowding (e.g., Evans and Kim, 2006; Marks and Choi, 2018; Maibom et al., 2021; Facchini, 2022). These papers show that the provision of medical procedures and care changes on crowded days compared to less crowded days. However, the effect of crowding on patient health outcomes is nil or very small. Our study contributes mostly to the second perspective, albeit focusing on productivity differences between different groups of professionals. Evidence comparing the productivity of different groups of professionals performing tasks contributing to the same outcome remains scarce, with the notable exception of Chan and Chen (2022), who causally show that nurses use more resources and achieve worse health outcomes than physicians on several measures in Veterans Health Administration Emergency Departments. Our study, which focuses on full teams in maternity units, adds to the literature by identifying differences in productivity across different team compositions. A major limitation of our setting is that we are not able to measure productivity variation *within each team* (unlike Chan and Chen, 2022), but only *across team compositions*. However, while most of the empirical literature comes from the US, we contribute to the scant literature measuring productivity in a tax-funded health care system characterised by public hospitals only, where the absence of residual claimants can strongly affect productivity.[1]

Second, this study is also related to the literature studying the relationship between volumes and outcomes (e.g. Gaynor et al., 2005). The general lesson from this literature is that high-volume hospitals provide better outcomes than smaller hospitals, and this may occur through different channels (Rachet-Jacquet et al., 2021). Higher volumes lead to better outcomes through improved technique or ability to detect and prevent complications, better collaboration, better work routines and coordination, learning-by-doing in medical skills or a better choice of treatment. We focus on the availability of a skill-mixed team of professionals as a crucial mechanism in explaining health outcomes in maternity units linking health outcomes with volume. A small body of literature that fits into this framework and bridges the literature on congestion concerns the study of the health effects of hospital consolidation and closure (Gaynor et al. 2012; Avdic et al., 2019, 2024). The closure of small hospitals, likely in remote or rural areas, reveals a trade-off between volume and quality, as opposed to patient travel distance and equity. Within this literature, interesting evidence on the effects of maternity unit closures in Sweden is provided by Avdic et al. (2024). Their main findings show that patients not directly affected by closures may suffer if capacity is not sufficiently expanded and hospitals become overcrowded. However, even patients exposed to closures may benefit if their new hospital is of higher quality. Our results complement this analysis with the identification of a possible mechanism, namely the role of the availability of medical staff. As the law specifies the available staff units according to the volume group to which the maternity unit belongs, we can relate the reduced availability of clinical staff in hospitals slightly below the legal thresholds to health outcomes by comparing them with those observed in hospitals with similar workloads but slightly above the thresholds.

The remainder of the paper is organized as follows: Section 2 provides essential background information, Section 3 introduces the data and the descriptive evidence, and Section 4 discusses the empirical strategy. We comment on the results in Section 5. Section 6 concludes, discussing the policy implications of our work.

2. Background

The Italian National Health Service provides universal coverage, largely free of charge at the point of delivery. According to the Constitution and the law 833/1978 (which established the Italian National Health Service), the central government guarantees appropriate funding for all regions to finance a set of essential services (the so-called Essential Levels of Care, *"Livelli Essenziali di Assistenza"*, or LEA), and the regional governments are responsible for the organization and the supply of health care services (e.g., Turati, 2013). The main funding sources are (mostly) national and regional taxes, supplemented by small pharmaceuticals and outpatient care co-payments which represent a minor share of total funding.

The current organization of maternal and newborn health services dates to the years 2000–2001 when the *"Progetto Obiettivo Materno Infantile"* (literally "Project targeting mothers and infants," Decrees of the Ministry of Health D.M. 24/04/2000 and D.M. 14/02/2001) was implemented. This legislation provides the main guidelines for healthy conception, pregnancy, birthing, and postnatal care. In addition, it determines that pregnancy care belongs to the set of Essential Levels of Care that must be guaranteed in all regions. Since this legislation was introduced, a few national and primarily regional implementation decrees have completed the general requirements of the national law.

In this paper, we concentrate on the Piedmont Region, a large and rich region in the North-Western part of the country with a population of around 4.3 million, comparable to countries like Croatia and Ireland. The Piedmont Region has fully implemented national guidelines since 2010 (D.G.R. n. 34–8769 of 12/5/2008 and the State-Regions Agreement 16/12/2010). The organization of perinatal care includes four phases: i) antenatal, ii) prenatal, iii) intrapartum, and iv) puerperium. The regional health system provides a wide range of free health services to pregnant women during each phase (e.g., Di Giacomo et al., 2022). These include obstetric visits, fetal ultrasound imaging, laboratory tests, antenatal education programs, and maternity hospitalization.

Our focus here is on the intrapartum phase. The regional network of

[1] Notable exceptions considering European countries are Avdic (2016), Friedrich and Hackmann (2017) and Daysal et al. (2019), which use data from Sweden, Denmark and the Netherlands, respectively. The first two studies do not focus on perinatal medical treatments and outcomes. The last article considers perinatal health outcomes, but only for a small part of the Dutch population, namely low-income mothers.

public hospitals for perinatal care follows a hub-and-spoke organizational design.[2] The network consists of six hub facilities (i.e., second-level maternity units) that provide a full range of health services, including neonatal intensive care, complemented by twenty-six secondary facilities (the spokes, i.e., first-level maternity units) that provide a more limited range of services. First-level maternity units treat women with a gestational age of more than 34 weeks, while second-level units treat women of any gestational age. Figure A1 in the Online Appendix shows the spatial distribution of maternity units across the regional territory. Mountains surround the regional territory on three sides (north, west, and south). Despite this geographical configuration, maternity units are fairly evenly distributed across the regional territory. In the event of maternal pathology or premature birth, the woman is referred to a second-level hospital by the regional maternity transport service.[3]

Maternity unit staffing, equipment, and supplies vary according to the hospital level (first- or second- level) and the annual number of births. Unsurprisingly, since capital and labor inputs are almost perfect complements in healthcare provision (e.g., Piacenza et al., 2010), the number of beds, delivery rooms, midwives per shift, obstetricians, gynecologists, and pediatric/neonatal staff depends on the annual number of births. Table A1 in the Online Appendix gives a more detailed account of the legal requirements. For instance, the law we utilise in this study specifies the minimum number of midwives per shift: for first-level maternity wards, no minimum requirements are defined if the annual number of births is below 500; the law requires two midwives per shift if the annual number of births is between 500 and 999, and at least three midwives for units with over 1000 births per year. Similar requirements, coherent with clinical staff requirements, are defined for capital inputs, such as the number of delivery rooms.

In addition, each maternity unit is assessed by a National Birth Path Committee (*Comitato Percorso Nascita nazionale*, CPNn) according to the operational, safety, and technological standards defined by the Ministry of Health (Ministero della Salute, 2010). Particular attention is paid to maternity units operating below the 500 births threshold, which may be allowed to operate if they serve geographical areas characterized by difficult access to healthcare services (like mountain areas, e.g., Perucca et al., 2019). In this case, the regional health departments must submit a reasoned request to the CPNn, which assesses whether the ward meets the necessary quality and safety requirements; if not, the CPNn defines actions to overcome the problems identified. After this process, the regional health department and the CPNn constantly monitor the maternity unit to verify its performance regarding maternal and neonatal health status.

3. Data

3.1. Data sources and sample definition

Our study is based on microdata from the Certificate of Delivery Assistance (*Certificato di Assistenza al Parto*, CeDAP) of the Piedmont region. The Certificate is mandatory and must be completed by the attending midwife or physician within ten days of delivery, but it is generally finalized soon after birth. The certificate contains epidemiological information on the mother's health status, sociodemographic characteristics, risk factors during pregnancy, obstetric procedures, and delivery methods. In addition, it includes any anomalies or congenital anomalies of the infant, causes of death (in case of stillbirth), information on the use of prenatal care services, etc.[4] The sources of all data in the certificate are medical records and official personal data, except for socio-economic information (marital status, educational level, and employment status), which is self-reported. The certificates also report data on the personnel attending the birth, specifying whether (or not) a midwife, a gynecologist, a pediatrician, and a nurse aide participated in the birth.

According to medical guidelines, a "team-based approach" to perinatal health care delivery involving midwives, obstetrician-gynecologist (OB/GYN) hospitalists, nurses, neonatologists, and other professionals (e.g., respiratory therapists, anesthesiologists, lactation consultants) is essential to improve the outcome of pregnancy (ACOG, 2017). Given the requirements defined by the Italian regulation, we aim to compare health outcomes at delivery for births attended by a "full team" with those obtained by a smaller team of professionals. Starting from the data, we define the "full team" as a team including at least a midwife, a gynecologist, and a pediatrician/neonatologist: this is the largest possible team observed in our sample. As an alternative, smaller teams are composed, for instance, only of a midwife or a midwife with a gynecologist (more details in Section 3.3).

We consider the years between 2011 and 2013, immediately following the end of the organizational process of maternity wards at the regional level required by the law. We stop our analysis in 2013 to avoid including the period, starting in 2014–2015, when some maternity wards were closed, which could have led to possible attrition problems if only more productive and large wards survived.

Our initial sample consists of 104,559 births in thirty-two public hospitals for which we know the mode of delivery, either vaginal (70 % of total births) or caesarean (the remaining 30 %). We then apply three sample restrictions. Online Appendix Table A2 details the number of observations dropped due to each restriction. First, we exclude women who do not live within the administrative boundaries of the Piedmont region. We then exclude second-level maternity units, i.e., the highly specialized hospitals with a neonatal intensive care unit that can treat all pregnancies, including those below 34 weeks gestation. Finally, we exclude observations without data on our control variables, including maternal, delivery, and hospital characteristics. The final sample consists of 55,840 observations. For the RDD analysis, we consider observations on either side of the two relevant thresholds of 500 and 1000 births per year, within windows of ± 250 births. This reduces the sample to maternity units within the range of 250–1249 births per year, corresponding to 35,067 observations.

Table A3 in the Online Appendix details the definition, and Table A4 reports the summary statistics of all outcome and control variables.

3.2. Outcome variables

We consider two sets of outcome variables: 1) the mode of delivery and 2) the health outcomes of the newborn and the mother at birth.

First, we consider the mode of delivery, an intermediate output of the production process in maternity wards, which is either a C-section or a vaginal birth. We construct a binary variable equal to one if the delivery was by C-section and zero if it was a vaginal birth. To identify the type of C-section or vaginal delivery more precisely, we also construct additional outcome variables as binary indicators. The binary variable *Emergency C-section* equals one for an unscheduled C-section, usually

[2] Private hospitals play a very minor role in deliveries – 98% of deliveries take place in public hospitals. The full cost of deliveries is borne by the public budget only in public hospitals.

[3] Gestational age is the number of weeks between the date of conception and the date of birth. A preterm birth (or premature birth) occurs before the 37th week of pregnancy. A term birth is between 37 and 41 weeks. A post-term (prolonged) pregnancy occurs when the baby is born at or after 41 weeks' gestation. If a woman presents to a first level maternity unit at <34 weeks' gestation, she will be transferred to a second level hospital. The pregnant woman is transferred to an emergency centre with specific expertise in pathology. Acute maternal transport reduces adverse maternal and neonatal outcomes.

[4] For further details, see Decree No. 349, dated 16 July 2001, of the Italian Ministry of Health.

required because of an immediate threat to the life of the woman or the fetus, and zero otherwise. *Planned C-section* equals one for a scheduled C-section, usually required for a specific medical diagnosis such as twin pregnancy, breech presentation, or communicable disease, and zero otherwise. *Assisted vaginal delivery* is equal to one if the vaginal delivery was performed using forceps or a vacuum device and zero otherwise; while *Uncomplicated vaginal delivery* is equal to one if the vaginal delivery was spontaneous and did not require the use of any device. A large literature identifies several determinants of C-sections.[5] Recent contributions focused on the role of workload (e.g., Facchini, 2022; Maibom et al., 2021) and management practices (e.g. La Forgia, 2022). Hospital management and workload are among the drivers of changes in the use of medical procedures. However, the impact on C-sections is mixed: on busy days, the number of C-sections can either increase (e.g., by diverting women to surgical procedures to reduce midwifery workload, Facchini, 2022) or decrease (e.g., by delaying induced labour or admitting women at later stages of labour, Maibom et al. 2021). We can go beyond workload and congestion explanations and add new evidence to this strand of the literature. We consider mode of delivery as an intermediate output, which can affect final health outcomes of the newborn and the mother, to test whether ward size affects decisions about mode of delivery at all. As the presence of different types of professionals in the delivery room is determined by ward size, we can also provide evidence on the role of different types of professionals in the delivery room on the choice between two potentially substitutable procedures, vaginal delivery and C-section.

Secondly, in our productivity regression model for maternity wards, we consider several newborn and maternal health status indicators at birth as final health outcomes (Avdic et al., 2024). As for newborn health status, we consider the probability of no need to be resuscitated, the Apgar score[6] measured after one minute and after five minutes (and their difference), and the likelihood of no meconium appearance. For the health status of mothers, we use the absence of lacerations of first, second, and third degree, the absence of episiotomy, and the length of the hospital stay.

Starting with newborn health status, the variable *No Need for Resuscitation* is a dichotomous variable that is equal to one if the newborn did not need any treatment with drugs, intubation, cardiac massage, or oxygen at birth, and zero if, on the contrary, the newborn needed a treatment. The Apgar scores are important measures to define how well the baby is doing outside the mother's womb: the measures are taken after one and five minutes from birth. *Apgar scores* are 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 three to eight (Figlio et al., 2014), and of school attainment, earnings, and social assistance receipt after age eighteen (Black et al., 2007; Oreopoulos et al., 2008). The Apgar score can be classified as follows: scores of 7 and above are considered normal, with 9 and 10 being perfect scores; scores from 4 to 6 are considered fairly low, while scores equal to or below 3 are critically low. A low one-minute score indicates that the newborn requires medical attention but does not necessarily correlate with long-term health problems, especially if the score improves at the five-minute threshold. We consider the two Apgar scores at one and five minutes as continuous variables and the absolute difference between the two scores (*ΔApgar*). We also define two dichotomous indicators for the probability of an Apgar score equal to or higher than 9 (a perfect score) for both the one and five-minute scores. Finally, *No meconium* is a dichotomous variable equal to one if no meconium appears during labor, zero otherwise. Meconium is a thick, sticky substance produced by the intestine of the newborn. Usually, it is released only after birth. If meconium appears during labor, fetal distress may occur as the newborn could aspirate the meconium during labor or delivery. Meconium aspiration syndrome may be a severe problem, resulting in pneumonia and the need for neonatal intensive care.

Considering maternal health outcomes, we use two dichotomous variables for the absence of perineal lacerations or tears, distinguishing first- from second or third-degree lacerations: 1) *No obstetric lacerations 1st degree,* and 2) *No obstetric lacerations 2nd-3rd degree*. Perineal lacerations are common during childbirth. A first-degree laceration involves only perineal skin, while a second or third-degree laceration involves some muscle tissues and usually requires stitches for optimal healing. We also define the variable *No episiotomy*, a dummy variable equal to one if an episiotomy was not needed. An episiotomy is a minor surgery cut to make it easier for the infant's head to pass through for delivery and to prevent complications or a vaginal tear. While episiotomy was a routine procedure in the past, the latest medical guidelines (e.g., ACOG, 2017) suggest that it should be done only for specific medical indications. Finally, we use a proxy for the presence of other maternal and infant complications by introducing an indicator variable *No other complications* equal to one if the number of days of hospitalization is equal to or less than two, corresponding to the standard hospital stay for a vaginal delivery.

We also define some additional outcomes to further understand the impact of medical staff on delivery. We do not observe postpartum health outcomes, but we complement our main health outcomes with a number of intrapartum or prepartum conditions. The *Breastfeeding* variable is a dummy equal to one if the mother was able to breastfeed the baby within two hours of giving birth. Early initiation of breastfeeding has positive effects on the health of newborns (reduced risk of infection and mortality), and it also facilitates the emotional bonding between mother and child (Hansen, 2016). The dummy *No Kristeller* is equal to one if the Kristeller maneuver - i.e., a controversial and risky procedure implying manual pressure on the fundus of the uterus towards the birth canal - was not performed. We also consider whether the placental expulsion was spontaneous (*Spontaneous afterbirth*), whether oxytocin or prostaglandins were (not) used to speed up the delivery (*No oxytocin* and *No prostaglandins*), and whether the rupture of the membranes was spontaneous (*No amniorrhexis*).

3.3. The "full team" indicator and the other controls

The main variable of interest in the productivity regression model for maternity wards is the "full team" binary indicator, which equals one if a team of professionals involving at least one midwife, a gynecologist, and a pediatrician (or a neonatologist) attended the birth. As already mentioned, we defined this variable based on our data. Table A5 in the Online Appendix shows all combinations of professionals attending the deliveries in our sample. The full team is the largest possible combination of professionals we observe. However, the likelihood of observing a full team varies across hospitals classified according to the annual number of births and by mode of delivery. Considering the two relevant thresholds of 500 births and 1000 births per year, the probability of observing a full team is monotonously increasing with the number of births for vaginal deliveries but not for C-sections. In particular, a full team is observed in 8.6 percent of vaginal deliveries in wards with 250–499 births, 24.3 percent of deliveries in units with 500–749 births, 48.9 percent of births in the 750–999 interval, and 52 percent of deliveries in the 1000–1249 interval. The corresponding percentages for C-sections are 43.3 for wards with 250–499 births, 82.6 for 500–749 births, 61.5 for 750–999 births, and 87.2 for wards with over 1000

[5] The factors traditionally investigated include financial incentives, that make C-sections more or less remunerative for physicians (i.e., supply-side factors or physician-induced demand, e.g., Gruber et al., 1999; Di Giacomo et al., 2017; Berta et al., 2020), patient preferences (i.e., demand-side factors, e.g., Amal-Garcia et al., 2022), a culture of defensive medicine to reduce the risk of litigation in problematic labour (e.g., Currie and MacLeod, 2008).

[6] The Apgar score is on a scale of 0 to 10. It is based on the sum of the scores ranging from 0 to 2 that are assigned to the following five criteria: appearance, pulse, grimace, activity and respiration. It is measured one and five minutes after delivery.

births.

We find different patterns when we look at the subcategories of vaginal and caesarean births. In general, a full team is more likely for C-sections than for vaginal deliveries. Looking into the two groups, we observe a full team more often for emergency than for planned C-sections in all hospital groups. Similarly, assisted vaginal births require a full team more often than uncomplicated spontaneous vaginal births. The most common team composition for uncomplicated vaginal deliveries in maternity units around the 500 births cut-off is midwife-gynecologist. For maternity units around the 1000 births cut-off, the most common team composition is the full team, followed by the midwife-gynecologist. Finally, for emergency and planned C-sections, the most common team composition is always the full team, followed by gynecologist-pediatrician. This evidence suggests that the production process for vaginal deliveries (especially uncomplicated ones) is likely to be different from that of C-sections.

After controlling for full team, we include a large set of mother, delivery, and hospital characteristics in all model specifications. The main purpose of including these variables is to control for various aspects of the prenatal environment that may influence newborn and maternal health outcomes at birth (Conti et al., 2020). The mother's characteristics controlled for are her socio-economic attributes, medical conditions, and lifestyle during the pregnancy. We include age, nationality, education level, employment status, marital status, whether the woman is at her first delivery, number of hospital admissions during the pregnancy, weight gain during pregnancy, smoking during pregnancy, and whether the woman experienced past abortions or miscarriages. Since proximity to the hospital may affect the labor stage on arrival (Card et al., 2023), all specifications include the distance from the mother's municipality to the nearest hospital. Over 75 % of the mothers in our sample give birth in the closest hospital.

We also include several delivery and infant characteristics: use of antibiotics during labor, neonatal head circumference, monitoring of the fetal heartbeat, gestational week, and whether the birth occurred on a weekend, public holiday or night shift. Women who are at increased risk of having a baby with Group B Streptococcus (group B strep, GBS) disease will be given antibiotics intravenously during labor. The antibiotics help protect the baby from infection. Neonatal head circumference is associated with infant well-being, and a large fetal head circumference may be associated with complicated labor. Fetal heart rate monitoring is quite routine, as an abnormal fetal heart rate may mean that the newborn is not getting enough oxygen or that there are other problems: the variable is introduced as a dummy variable equal to one if there was no fetal heart rate monitoring during labor. Finally, we control for the gestational age of the newborn in weeks and the timing of the birth by including two dummy variables. The first dummy is for festivities and is equal to one if the birth took place on a weekend or a public holiday. The second dummy variable is for nights and is equal to one if the birth took place during the night (from 00:00 to 08:00).

Finally, we introduce the ratio of femoral neck fractures treated with surgery within two days at the hospital level as a proxy for the hospital's effectiveness: the higher the ratio, the more effective the hospital is from a clinical and organizational point of view.[7] This measure comes from the *"Programma Nazionale Valutazione Esiti"* (P.N.E., literally "National Program for the Evaluation of Outcomes"), a program financed by the Italian Ministry of Health that collects a wide range of information on Italian public and private hospitals to support clinical and organizational audit programs.

[7] Other indices commonly used as proxies for hospital efficiency relate to 30-day outcomes after acute myocardial infarction. We use the treatment of fractures of the femoral neck because data are available for a larger number of hospitals, as many hospitals may not have a cardiac unit.

4. Empirical strategy

We estimate the following specification for the productivity regression model:

$$Y_{iht} = \beta_1 FT_{iht} + \sum_{k=1}^{2} \rho_k B_{ht}^k + \boldsymbol{M}'_{iht}\boldsymbol{\beta}_2 + \boldsymbol{D}'_{iht}\boldsymbol{\beta}_3 + \boldsymbol{H}'_{ht}\boldsymbol{\beta}_4 + \theta_p + \theta_m + \theta_t + \varepsilon_{iht} \tag{1}$$

where Y_{iht} is, in turn, the "method of delivery" for mother i, or the "health status" of the newborn and mother i at birth, discharged from hospital h, in year t. B_{ht} is the number of births in hospital h in year t, entered linearly and squared. They provide a measure of current hospital volume and allow us to account for economies of scale or learning-by-doing effects (Gaynor et al., 2005; Rachet-Jacquet et al., 2021; Avdic et al., 2024). $\boldsymbol{M}$, $\boldsymbol{D}$, $\boldsymbol{H}$ are matrices of mother, delivery, and hospital characteristics that represent risk adjustment variables for the outcomes at birth.[8] FT is the binary indicator for the presence of a full team of professionals assisting the newborn or mother i during the delivery. The term θ_p is an i.i.d. Local Health Authority component. In particular, the Piedmont region is divided into fourteen areas managed by Local Health Authorities (LHA or *Aziende Sanitarie Locali*), and each authority is responsible for providing health services in its specific geographical area.[9] The terms θ_m and θ_t are i.i.d. month and year components, respectively. These error components are introduced to parameterize possible correlations in health status within the district of the local health authority and within time. The remaining error component ε_{iht} is specific to newborn or mother i. The standard errors are clustered at the hospital, working shift, and weekday levels in all specifications to account for the likely correlation of residuals within these categories.

The FT coefficient β_1 is the parameter of primary interest. Eq. (1) describes the likelihood of a particular mode of delivery and the average potential health outcomes for newborns and mothers under alternative health team assignments, controlling for any effects of maternal, delivery, and hospital characteristics collected in matrices $\boldsymbol{M}$, $\boldsymbol{D}$, and $\boldsymbol{H}$. However, estimating the causal impact of the team on our outcomes is challenging. One main reason is that the hospital can adjust labor inputs according to the patient's needs. In particular, the presence of a full team will depend on two main factors. First, the characteristics of the mother and the birth may determine the composition of the team. These characteristics include the type of pregnancy (singleton or multiple), the week of pregnancy (term or preterm), and any medical or obstetric complications that may arise during labor. Second, the characteristics of the hospital and its internal organization may limit the composition of the team, e.g., the time of delivery (night, weekend, or day shifts) and, more generally, the availability of hospital staff. Since a full team is not randomly assigned, it will likely correlate with the error components in Eq. (1). OLS estimates of the full team coefficient β_1 from Eq. (1) are likely to be biased if hospitals choose different observable teams for patients who differ in their unobserved latent characteristics. In addition, the selection of mothers into hospital facilities may not be random. A woman and her partner can choose the hospital where she gives birth. Usually, parents choose the nearest maternity hospital. However, the choice is also influenced by the presence of obstetric or medical

[8] In all specifications, matrix **M** includes the mother's age, nationality, education level, employment status, marital status, whether the woman is at the first delivery, number of hospitalizations during pregnancy, weight gain during pregnancy, smoking during pregnancy, whether the woman experienced previous abortions and miscarriages, and the distance (in Km) between the mother's municipality and the nearest hospital. Matrix **D** accounts for the use of antibiotics during labor, neonatal head circumference, fetal heart rate monitoring, gestational week, and whether the birth occurred on a weekend or at night. Matrix **H** includes the proportion of femoral neck fractures treated within two days. All definitions are in Table A3 in the Appendix.

[9] Six of the fourteen local health authorities in Piedmont have only one maternity ward.

complications (e.g., high-risk pregnancy, premature birth, diabetes, etc.) or the desire to give birth in the hospital where the doctor/midwife attending the pregnant woman has admitting privileges.

We comprehensively address endogeneity concerns for potential bias due to the endogenous selection of the team and the endogenous selection of patients into hospitals by exploiting the Italian legislation that defines different thresholds for staff and equipment requirements, that change discontinuously at each threshold. To obtain causal estimates of our parameter of interest, we experiment with two complementary methodologies: we apply the 2SLS-IV strategy on the entire available sample, and the RDD strategy on maternity units around the thresholds.

The 2SLS-IV strategy builds on the research design in Angrist and Lavy (1999), often called the Maimonides' Rule, with which the authors exploit class size cut-offs imposed by a rule in Israel to estimate the impact of class size on educational achievement. Similarly, here, we exploit cut-offs defined by the law to determine maternity care units' staffing, equipment, and supplies according to the yearly number of births to study the impact of staff inputs on health outcomes. Staff, and availability of other resources discontinuously increase at each threshold since input requirements change according to the average volume of deliveries in the previous three years: below 500 deliveries, in the interval 500–999, or larger than 1000 (see Table A1). This allows us to define three instrumental indicator variables $\boldsymbol{I}[\boldsymbol{MaternSize}]$ for full teams, based on intervals in the number of births in a specific hospital h, in the previous years: the first variable indicates hospitals with fewer than 500 births, the second variable indicates hospitals in the interval of 500–999 deliveries, and the third indicates hospitals with >1000 births.

An instrument must fulfil some key assumptions to be valid. First, it must be related to the endogenous variable. In our case, the legal requirements imply that the regressor of interest (FT) is partly determined by a known discontinuous function of an observed covariate (the number of births in the previous years). We present a large set of results from the first stage (on the full sample and on some subsamples of mode of delivery) and the corresponding F-statistics on the excluded instrumental variables to check the relevance of the instrument. The first-stage equation for the 2SLS-IV model in (1) is the following:

$$FT_{iht} = \boldsymbol{I}[\boldsymbol{MaternSize}]_h{}'\gamma_1 + \sum_{k=1}^{2}\pi_k B_{ht}^k + \boldsymbol{M}'_{iht}\boldsymbol{\gamma}_2 + \boldsymbol{D}'_{iht}\boldsymbol{\gamma}_3 + \boldsymbol{H}'_{ht}\boldsymbol{\gamma}_4 + \lambda_p + \lambda_m + \lambda_t + \eta_{iht} \quad (2)$$

As in Eq. (1), B_{ht} is the number of births in hospital h in year t, while $\boldsymbol{M}$, $\boldsymbol{D}$, $\boldsymbol{H}$ are matrices of mother, delivery, and hospital characteristics, while λ_p, λ_m and λ_t are local health authorities, monthly and year-fixed effects, respectively. We repeat the analysis for the full sample of all births and some subsamples based on the method of delivery (vaginal births and C-sections).

Second, the instruments must not share any unobservable/unmeasured common causes with the outcome. We assume that any other effects of the number of births on health outcomes are adequately controlled by the terms included in Eq. (1) and "partialled out" of the instrument by the variables in $\boldsymbol{M}$, $\boldsymbol{D}$, $\boldsymbol{H}$, and the fixed effects.

Third, the instrumental variable must not affect the outcome other than through the endogenous variable. We discuss the plausibility of the exclusion restriction and try to identify possible violations by exploiting the discontinuity in the assignment mechanism as in Angrist and Lavy (1999). In addition, we present Sargan-Hansen test results for the validity of over-identifying restrictions. In our context, the main threat to identification is the fact that crossing the threshold may affect health not only through staffing but also through the availability of additional resources and equipment. We do not have data to provide direct evidence of differences in the equipment and the number of delivery rooms and beds available to the wards in our sample. However, we provide indirect evidence on both issues in Section 5.1. In particular, we exploit a set of proxies for equipment endowment during labour and delivery and for the number of delivery rooms and beds to argue that the only effect of crossing the threshold comes from team composition.

A final identifying assumption is that parents do not selectively exploit the rule to deliver in hospitals with larger maternity wards. Selective manipulation could occur, for example, if more educated parents choose hospitals with many births, knowing that this will result in a full team of professionals attending the delivery. In practice, however, parents do not participate in the decision-making process that leads to the actual team of professionals. First, we observe that over 75 % of mothers in our sample deliver at the closest hospital, which is consistent with other studies that similarly find that mothers choose to deliver at the closest hospital (e.g., Phibbs et al., 1993; Currie and MacLeod, 2017; Card et al., 2023). Second, we test for the absence of ad hoc manipulation around the cut-offs and introduce a specification that relies on a regression discontinuity design (RDD). We focus on maternity wards around the 500 and the 1000 number of births thresholds, and we employ a Fuzzy RDD, where the discontinuities induced by having a number of births above the 500 or the 1000 births thresholds are used as an instrument for the full health care team. Within this framework, we limit the analysis to hospitals whose number of births is in the bandwidth ±250 births around the two cut-off points.[10] We thus assume that hospitals located right above or below the threshold are very similar in workload. However, those above are required to have more staff units available and, coherently, additional beds and equipment.

We then estimate the following RDD reduced form model for the two cut-off points considered (either $B^* = 500$ or $B^* = 1000$):

$$Y_{iht} = \pi_1 A_{ht} + \sum_{k=1}^{2}\sigma_k q_{ht}^k + \boldsymbol{M}'_{iht}\boldsymbol{\pi}_2 + \boldsymbol{D}'_{iht}\boldsymbol{\pi}_3 + \boldsymbol{H}'_{ht}\boldsymbol{\pi}_4 + \eta_p + \eta_m + \eta_t + u_{iht} \quad (3)$$

where, like in Eq. (1), Y_{iht} is the delivery method or health status at birth of the infant/mother i, in hospital h at time t, while the dummy A_{ht} is equal to one if hospital h at time t is above the considered threshold and zero otherwise. In the specification, we also include the distance to the cut-off point q (and its square), defined as $q_{ht} = (B_{ht} - B^*)$, where B_{ht} is the number of births in hospital h, at time t, and B^* is alternatively equal to 500 or 1000, the cut-off points defined by the law. We also include the complete set of mother, delivery, and hospital characteristics ($\boldsymbol{M}$, $\boldsymbol{D}$, $\boldsymbol{H}$) while η_p, η_m, and η_t are local health authorities and month- and year-fixed effects, respectively.

We alternatively estimate the specifications in (3) on three different samples: (i) the sample of births occurring in all hospitals within the bandwidths of ±250 births around the 500 and 1000 thresholds, (ii) the sample of births occurring only in hospitals around the 500 threshold, and (iii) the sample of births occurring only in hospitals around the 1000 threshold. Also for the RDD approach, we estimate a first stage specification in which the full team *FT* variable is regressed on the discontinuity in the availability of staff induced by the law (A_{ht} dummy for being above the threshold):

$$FT_{iht} = \delta_1 A_{ht} + \sum_{k=1}^{2}\tau_k q_{ht}^k + \boldsymbol{M}'_{iht}\boldsymbol{\delta}_2 + \boldsymbol{D}'_{iht}\boldsymbol{\delta}_3 + \boldsymbol{H}'_{ht}\boldsymbol{\delta}_4 + \mu_p + \mu_m + \mu_t + \xi_{iht} \quad (4)$$

To examine the heterogeneity of the results, we repeat the analysis on the full sample of all births, on subsamples based on the method of delivery (vaginal and caesarean sections), and on subsamples based on other characteristics of the birth (time of day, weekend, native or non-native mother).

[10] We also experiment with different bandwidths around the cut-off points (±100; ±150; ±200) and results do not qualitatively change, even if we find larger standard errors when reducing the bandwidth below 100.

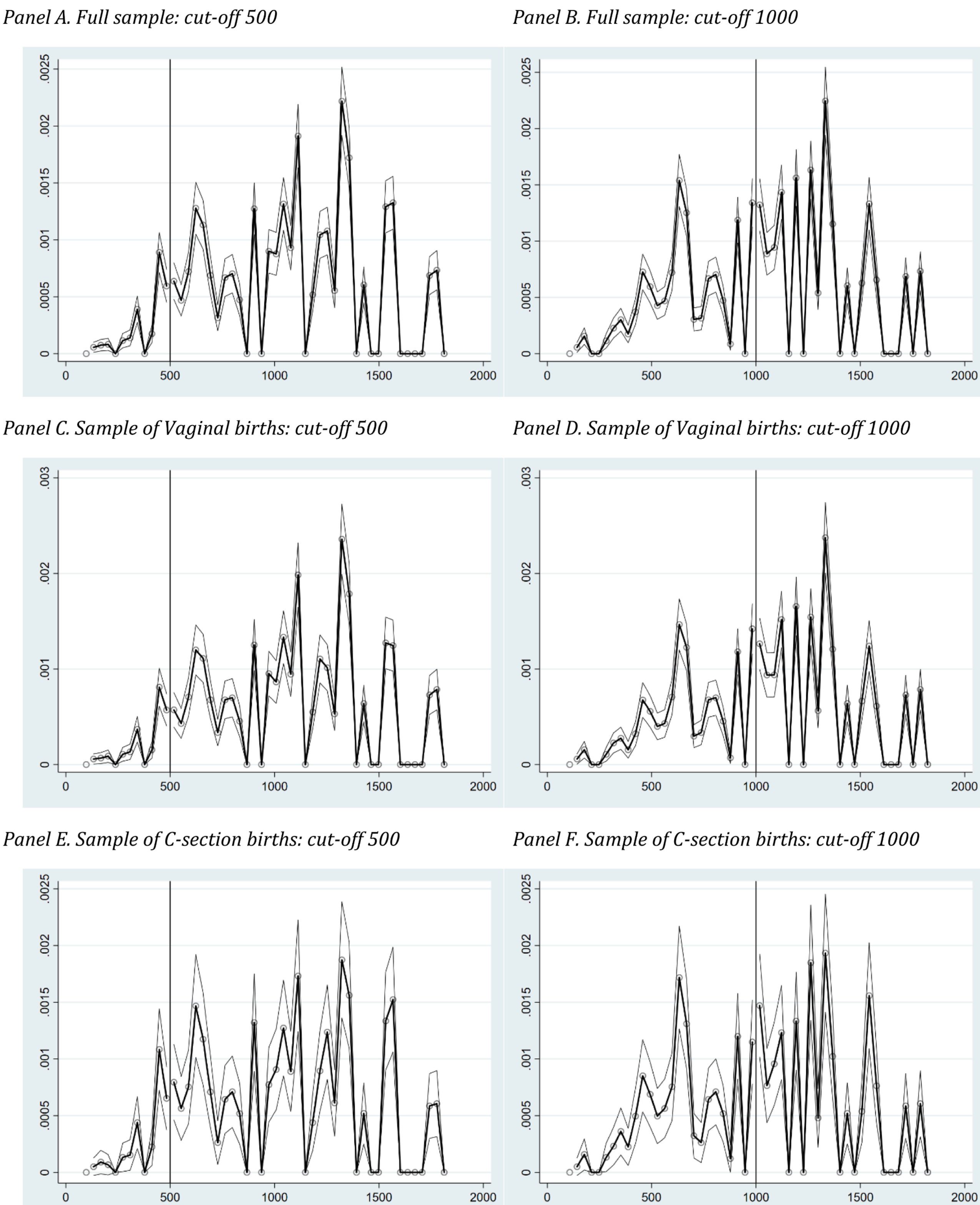

Fig. 1. Births distribution across thresholds. Panel A. Full sample: cut-off 500, Panel B. Full sample: cut-off 1000. Panel C. Sample of Vaginal births: cut-off 500. Panel D. Sample of Vaginal births: cut-off 1000. Panel E. Sample of C-section births: cut-off 500. Panel F. Sample of C-section births: cut-off 1000.
Notes: Each panel represents the estimated discontinuous density functions for the yearly number of births at hospital level (McCrary, 2008). The binsize and bandwidth were chosen using the automatic procedure in DCdensity for Stata (McCrary, 2008). The solid thick line displays the densities of the McCrary (2008) estimator and the solid thin line the associated 95 confidence interval. On the left panels the cut-off point is 500 births per year, while on the right panels the cut-off point is 1000 births. The number of observations is: 55,840 for Panels A and B; 39,839 for Panels C and D; 16,001 for Panels E and F. A *t*-test of the null hypothesis of continuity fails to reject for each panel at any significance level, i.e., the density estimates are consistent with continuity at the thresholds.

Table 1
Overidentification tests.

Panel A. Medical conditions and socio-economic characteristics of the mother

	Full sample		*Vaginal deliveries sample*		*C-Section sample*	
	500 births (1)	1000 births (2)	500 births (3)	1000 births (4)	500 births (5)	1000 births (6)
Age	0.4276*	−0.1790	0.2830	−0.2204	0.7218**	−0.4445
	(0.238)	(0.271)	(0.253)	(0.274)	(0.329)	(0.446)
Native	0.0483**	−0.0486***	0.0518**	−0.0443**	0.0413	−0.0661**
	(0.019)	(0.017)	(0.025)	(0.020)	(0.031)	(0.025)
Primary education	−0.0096**	0.0085**	−0.0077	0.0109**	−0.0102	0.0030
	(0.004)	(0.004)	(0.007)	(0.005)	(0.006)	(0.006)
Secondary education	−0.0153	−0.0292**	−0.0175	−0.0262	−0.0116	−0.0367*
	(0.013)	(0.013)	(0.018)	(0.017)	(0.027)	(0.020)
Tertiary education	0.0248*	0.0207	0.0253	0.0153	0.0218	0.0337*
	(0.014)	(0.013)	(0.019)	(0.017)	(0.026)	(0.020)
Employed	0.0337*	−0.0560***	0.0342	−0.0469**	0.0292	−0.0936***
	(0.018)	(0.016)	(0.022)	(0.021)	(0.030)	(0.026)
Married	−0.0327**	−0.0726***	−0.0320**	−0.0803***	−0.0322	−0.0455
	(0.014)	(0.021)	(0.015)	(0.028)	(0.032)	(0.042)
First child	0.0284	−0.0026	0.0611***	0.0006	−0.0317	−0.0183
	(0.020)	(0.022)	(0.020)	(0.026)	(0.065)	(0.084)
Hospital admissions during pregnancy	−0.0065	−0.0034	−0.0071	−0.0080	−0.0047	0.0073
	(0.025)	(0.005)	(0.024)	(0.005)	(0.030)	(0.010)
Weigth gain during pregnancy	2.2389***	−3.3116***	2.2808***	−3.6053***	2.1716**	−2.5618*
	(0.749)	(0.982)	(0.645)	(0.932)	(1.052)	(1.408)
Smoking during pregnancy	0.0067	−0.0348***	0.0082	−0.0324**	0.0033	−0.0406
	(0.010)	(0.012)	(0.013)	(0.013)	(0.013)	(0.027)
Previous abortion	0.0042	−0.0227**	0.0081	−0.0235*	0.0028	−0.0189
	(0.011)	(0.011)	(0.014)	(0.012)	(0.016)	(0.021)
Previous miscarriage	0.0284	−0.0186	0.0239	−0.0179	0.0397	−0.0210
	(0.021)	(0.021)	(0.023)	(0.018)	(0.031)	(0.043)
Km to closest Hospital	0.0051	1.0612***	−0.1530	1.0790***	0.3590	1.0021
	(0.489)	(0.397)	(0.472)	(0.373)	(0.639)	(0.646)
Delivery week	0.0242	−0.2371**	0.1433**	−0.1798***	−0.2389**	−0.2692
	(0.085)	(0.099)	(0.067)	(0.059)	(0.115)	(0.174)

Panel B. Delivery Characteristics

	Full sample		*Vaginal deliveries sample*		*C-Section sample*	
	500 births (1)	1000 births (2)	500 births (3)	1000 births (4)	500 births (5)	1000 births (6)
Antibiotics during labor	−0.0002	−0.0577***	0.0135	−0.0672***	−0.0258	−0.0077
	(0.021)	(0.016)	(0.022)	(0.015)	(0.022)	(0.029)
Newborn head circumference	0.0603	−0.1017	0.0795	−0.0182	0.0645	−0.3616*
	(0.100)	(0.106)	(0.127)	(0.116)	(0.190)	(0.194)
No fetal heartbeat monitoring	0.0181	−0.0108	0.0030	−0.0132	0.0510	−0.0651
	(0.020)	(0.020)	(0.013)	(0.017)	(0.032)	(0.065)
Week-end	−0.0328	−0.0056	−0.0330	−0.0036	−0.0402	0.0098
	(0.132)	(0.182)	(0.142)	(0.191)	(0.116)	(0.157)
Night shift	−0.0369	0.0110	−0.0336	0.0365	−0.0463	−0.0380
	(0.192)	(0.253)	(0.197)	(0.256)	(0.187)	(0.221)

Panel C. Hospital Characteristics.

	Full sample		*Vaginal deliveries sample*		*C-Section sample*	
	500 births (1)	1000 births (2)	500 births (3)	1000 births (4)	500 births (5)	1000 births (6)
Femoral fractures 48hrs	7.4305	20.4524	–	–	–	–
	(10.250)	(12.012)				
Relative num. shift deliveries	0.0759***	−0.0468	0.0621**	−0.0402	0.1034***	−0.0638
	(0.027)	(0.042)	(0.027)	(0.046)	(0.033)	(0.050)
Relative num. shift deliveries above 50 %	0.0150	−0.0170	0.0110	−0.0200	0.0230**	−0.0066
	(0.011)	(0.018)	(0.016)	(0.021)	(0.009)	(0.020)
Relative num. hourly deliveries	0.0138	−0.0197	0.0114	−0.0205	0.0176*	−0.0129
	(0.010)	(0.015)	(0.014)	(0.019)	(0.010)	(0.015)
Relative num. hourly deliveries above 50 %	0.0142	−0.0133	0.0160	−0.0116	0.0060	−0.0132
	(0.014)	(0.028)	(0.016)	(0.031)	(0.016)	(0.027)

Notes: In each row, we estimate an OLS equation where each single mother, delivery, or hospital characteristic is a placebo outcome regressed on the dummy above threshold, equal to one if the hospital has >500 births (odd columns) or 1000 births (even columns) yearly births. In all columns, we include deliveries in maternity units that record yearly births within the ±250 bandwidth around the thresholds of 500 and 1000, respectively. We include the difference (and its square) between the total number of births at the hospital-year level and the closest threshold (500 or 1000) in all specifications. The table reports only the coefficient for the above threshold dummy variable. The number of observations for the full sample is: 14,281 for the 500 births cutoff (column 1) and 20,786 for the 1000 cutoff (column 2). The number of observations for the vaginal deliveries sample is: 9754 for the 500 births cutoff (column 3) and 15,064 for the 1000 cutoff (column 4). The number of observations for the C-section sample is: 4527 for the 500 births cutoff (column 5) and 5722 for the 1000 cutoff (column 6). The variable *Femoral fractures 48* hrs is available at hospital/year level and the number of observations for that variable are 24 in column 1, and 45 in column 2. Standard errors (in parentheses) are clustered at the hospital, shift, and weekday level. Online Appendix Table A3 reports the definition of all variables. Significance levels: *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

5. Results

5.1. Threats to identification

Before presenting our results, we discuss possible threats to the identification strategy by investigating *ad hoc* manipulation around the cut-offs as an important concern for the identification strategy based on the Maimonides rules. We present two types of evidence: sorting of hospitals around the cut-offs and balancing tests for the comparability of maternal, delivery and hospital characteristics around the cut-offs.

Fig. 1 shows the estimated discontinuous density functions for the annual number of births at the hospital level around the two thresholds based on the number of births, according to McCrary (2008). We find no statistically significant evidence of manipulation. A *t*-test of the null hypothesis of continuity is not rejected for any sample (full sample, vaginal only, and C-sections only) at the usual levels of significance, i.e., the density estimates are consistent with continuity at the two thresholds of 500 and 1000. This evidence supports the absence of strategic manipulation of hospitals or maternity units into different size groups.

Next, we test for changes in the observable characteristics of women, deliveries, and hospitals around the cut-offs, checking for their smoothness around the two thresholds. Table 1 shows a full set of overidentification tests for each sample (full sample, vaginal only, and C-sections only) and for each group of characteristics (maternal, delivery, and hospital characteristics in panels A, B, and C, respectively). We estimate local-linear regressions using the set of mothers', delivery, and hospitals' characteristics as dependent variables.

As for maternal characteristics (Panel A of Table 1), results suggest that observations just below the 500-births cut-off are similar to those just above (column 1 of Table 1). Similarly, the observed characteristics are remarkably comparable around the threshold of 1000 births (column 2 of Table 1). However, there are some variables for which we find statistically significant differences. Women giving birth in hospitals just above the 500 threshold are older, more educated, more likely to be native, and employed, but less likely to be married when compared to women giving birth in hospitals below the 500 threshold. They are also more likely to have gained weight during pregnancy. On the contrary, women in hospitals above the threshold of 1000 births are less educated, less likely to be native and to be employed and married, and they have a lower gestational age. They also live further away from the hospital, gain less weight during pregnancy, are less likely to smoke, and are less likely to have had a previous abortion than women in hospitals below the threshold of 1000 births. Overall, the evidence suggests that there are differences in some characteristics of women across the thresholds. These characteristics are a mix of demographic and clinical characteristics around the 500 threshold (e.g., age, nationality, higher weight gain), whereas they are more medical/clinical characteristics around the 1000 threshold (e.g., lower weight gain coupled with lower gestational week, previous abortions, women living further from the nearest hospital). When splitting the sample according to the delivery method, we find that some significant results are exclusively due to vaginal deliveries (e.g., the marital status, smoking, distance to the nearest hospital), while others occur in both subsamples (e.g., nationality, weight gain).

Panels B and C of Table 1 show results for some delivery and hospital characteristics. First, we consider the use of antibiotics during labour, newborn head circumference, and fetal heart rate monitoring. We find no difference in newborn head circumference across the thresholds, while there is a lower use of antibiotics during labour above the 1000 threshold for vaginal births. As for fetal heart rate monitoring, we find no statistically significant coefficients: since this procedure requires a specific portable device, this result can be seen as indirect (albeit rough) evidence that there is no difference in the actual use of equipment across the cut-offs.

We then look at the availability of delivery rooms and beds, which could be proxied by a number of measures of crowding: deliveries during weekends and nights, the relative number of additional births in the same working shift[11] in the same hospital, and the relative number of additional births in the same hour in the same hospital. The relative number of additional deliveries is constructed as in Avdic et al. (2024). This measure is the ratio of actual deliveries during the same working shift in the same hospital (or in the same hour in the same hospital) to the expected number of births in each hospital and shift (hour).[12] We also introduce the binary version of the relative number of births, which is equal to one if the number of actual births is 50 per cent higher than the number of expected births, calculated at the working shift or hour level. We do not find significant differences between the thresholds for weekend and night deliveries. Similarly, we find no significant difference in the relative number of births per hour. We do find a significantly higher number of actual deliveries compared to the expected number of births in the same working shift above the 500 threshold, which implies that crossing the threshold tends to increase crowding.

Overall, the differences in the proxy variables for equipment and other resources across the thresholds are not significant or of small magnitude and opposite sign to what would be expected. Hence, we can cautiously rule out that the availability of more equipment and other resources (e.g. more delivery rooms) for hospitals above the thresholds allows for less crowding and thus better health outcomes. Therefore, it is reasonable to assume that having more resources, other than staff, does not affect the choice of mode of delivery and neonatal and maternal health outcomes.

Finally, we observe no significant differences in the proportion of femoral neck fractures surgically treated within two days at hospital level, suggesting no differences in the quality of healthcare treatments around the thresholds (OECD, 2018).

Figure A2 in the appendix plots the same observed characteristics of women and births against the annual number of births at hospital year level around the cut-offs. The different panels in Figure A2 show that most of the observed characteristics are smoothly distributed around the cut-offs, while a few show some relatively small discrete jumps. To account for these differences, we include all covariates describing maternal and birth characteristics in our analysis.

5.2. First stage estimates

We now discuss the importance of the law requirements on the probability of observing a full team. We first provide a graphical representation of the relationship between the instruments and the endogenous variable of the model, namely the full team of professionals attending the delivery. Fig. 2 plots the proportion of deliveries attended by a full team against the annual number of deliveries in a particular hospital and year. There is clear evidence of a positive relationship between the probability of observing a full team and the number of deliveries, with some discrete jumps around the 500 and the 1000 thresholds in the full sample (Panel A of Fig. 2) and also in the two subsamples of vaginal and C-section deliveries (Panels B and C of Fig. 2).

Table 2 reports first stage estimates of Eq. (2) in column 1, and estimates of Eq. (4) in columns 2, 3, and 4. Panel A considers all births, while Panels B and C report results for the subsamples of vaginal births and C-sections, respectively. In all specifications, the dependent variable is *FT*, a binary indicator for the presence of a full team during the delivery.

In column 1 of Table 2 (Panels A, B, and C), we include the two instruments: '*500–999*', which is equal to one if the maternity unit where the childbirth occurred has a number of yearly births in the intervals

[11] We use three working shifts: 6am to 2pm; 2pm to 10pm, 10pm to 6am.

[12] The expected number of deliveries is obtained from an OLS regression of actual births on hospital fixed effects and a set of time fixed effects: year, month, weekday, working shift or hour. We then use these estimates to obtain the expected number of births for each hospital and working shift or hour.

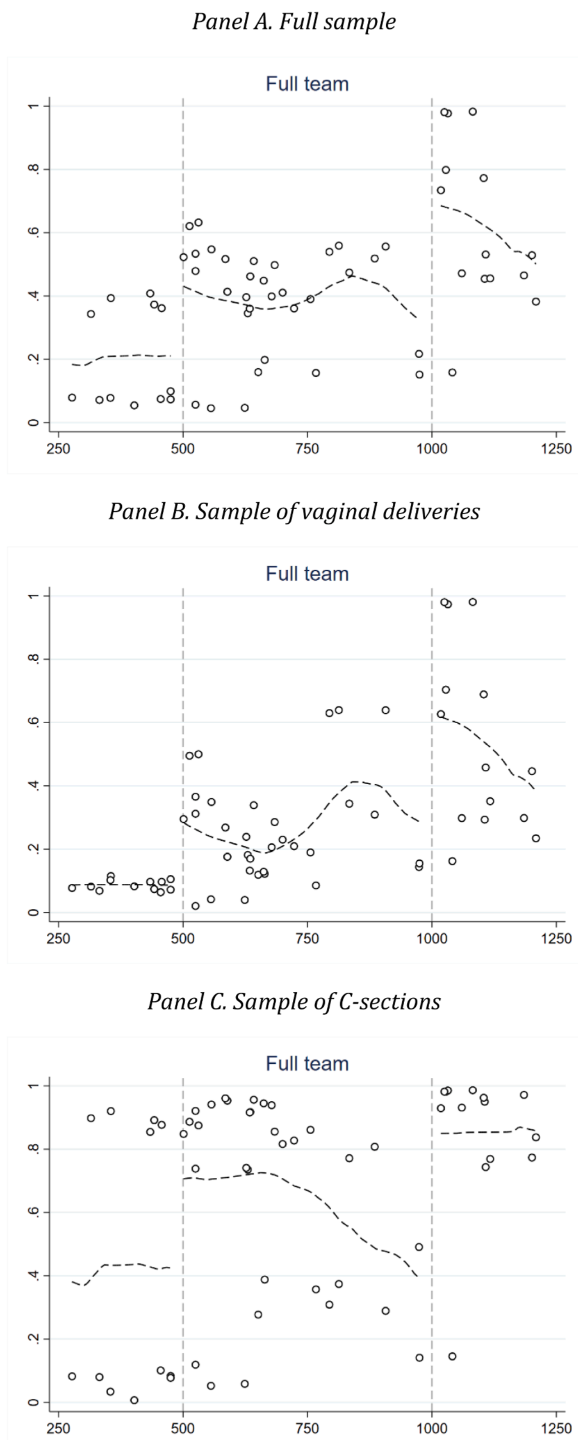

Fig. 2. Presence of full team during delivery. Panel A. Full sample. Panel B. Sample of vaginal deliveries. Panel C. Sample of C-sections.
Notes: Each panel represents the annual share of births attended by a full team at hospital level. The y-axis measures the proportion of total (in panel A, total number of observations 35,067), vaginal (in panel B, total number of observations 24,818) and caesarean (in panel C, total number of observations 10,249) births attended by a full team. The x-axis measures the number of births per year. Each point is a hospital-year. The two thresholds are at 500 and 1000 annual births.

Table 2
First stage estimates.

Panel A. Full sample				
	2SLS Sample (1)	*Discontinuity sample* (2)	*Sample 250–749 births* (3)	*Sample 750–1249 births* (4)
500–999	0.3490***			
	(0.048)			
Over 1000	0.7275***			
	(0.091)			
Above thresholds		0.2950***		
		(0.052)		
500–749			0.2256***	
			(0.038)	
1000–1249				0.1802***
				(0.056)
Mean of Full team	0.500	0.491	0.363	0.570
SD of Full team	0.500	0.500	0.481	0.495
F-Stat	35.63			
Observations	55,840	35,067	14,281	20,786
Panel B. Vaginal deliveries sample				
	2SLS Sample (1)	*Discontinuity sample* (2)	*Sample 250–749 births* (3)	*Sample 750–1249 births* (4)
500–999	0.3288***			
	(0.065)			
Over 1000	0.7407***			
	(0.115)			
Above thresholds		0.3146***		
		(0.044)		
500–749			0.3772***	
			(0.039)	
1000–1249				0.1424***
Mean of Full team	0.380	0.383	0.198	0.489
SD of Full team	0.485	0.486	0.398	0.500
F-Stat	20.87			
Observations	39,839	24,818	9754	15,064
Panel C. C-sections sample				
	2SLS sample (1)	*Discontinuity sample* (2)	*Sample 250–749 births* (3)	*Sample 750–1249 births* (4)
500–999	0.3465***			
	(0.082)			
Over 1000	0.5914***			
	(0.105)			
Above thresholds		0.2845***		
		(0.084)		
500–749			0.0219	
			(0.060)	
1000–1249				0.2980***
				(0.088)
Mean of Full team	0.796	0.750	0.710	0.778
SD of Full team	0.403	0.433	0.454	0.415
F-stat	16.20			
Observations	16,001	10,249	4527	5722

Notes: Column 1 reports the estimation of Eq. (2), while columns 2 to 4 the results for Eq. (4) where the dependent variable is the endogenous variable *Full Team FT*. The estimates in Panel A are based on full sample of all deliveries, while in Panels B and C we restrict the samples to vaginal and C-section births respectively. The dummy variables '500–999' and 'Over 1000' are one if the birth took place in a maternity unit with an annual number of births in the corresponding interval, and zero otherwise. The estimates in columns 2 to 4 are based on the RDD samples: the discontinuity sample in column (2), including all deliveries in maternity units recording annual births within the ±250 range around the thresholds of 500 and 1000; all deliveries in maternity units recording annual births within the ±250 range around the threshold of 500 (250–749) in column (3); all deliveries in maternity units recording annual births within the ±250 range around the threshold of 1000 (750–1249) in column (4). The dummy variables 'Above thresholds'; '500–749' and '1000–1249' equal one if the delivery took place in a maternity unit with an annual number of births in the corresponding interval, and zero otherwise. In all specifications, we also include mother's characteristics (age, nationality, level of education, employment status, mother's marital status, whether this is the woman's first birth, week of delivery, whether the woman had any hospitalisation during pregnancy, weight gain during pregnancy, smoking, experience of previous abortion and previous miscarriage, distance to nearest hospital) and delivery characteristics (newborn head circumference, fetal heart rate monitoring, weekend delivery, night shift delivery, antibiotics during labour), hospital characteristics (proportion of femoral neck fractures treated within two days), local health authority, month and year fixed effects. *Mean of Y* and *SD of Y* report the dependent variable's sample mean and standard deviation, respectively. *F-stat* is the F-statistics on the external instrumental variables in the first stage regression. Standard errors are clustered at hospital, shift and weekday level. Significance levels: *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

500–999, and zero otherwise; and '*Over 1000*' which is equal to one if the maternity unit has a number of yearly births larger than 1000, and zero otherwise. Estimated coefficients are always positive and significantly different from zero in all specifications. For the full sample of all births, being in a hospital with a number of births falling in the interval 500–999 increases the probability of having a full team of specialists by 34.9 percentage points while being in a hospital with a number of births above 1000 increases the probability of having a full team of specialists by 72.8 percentage points, with respect to the omitted category (hospitals with <500 births per year). Results are similar for the sub-samples of vaginal and C-section deliveries. Tests for under- and weak-identification suggest that the first stage is very precise, and the instruments are relevant.

In columns 2, 3, and 4 of Table 2 (Panels A, B, and C), we also present the first stage results for the smaller RDD samples around the 500 and 1000 cut-offs. In column 2 we consider the 'overlapped' discontinuity sample of maternity units in the intervals 250–749 and 750–1249 births per year, near the 500 and 1000 cut-offs. In columns 3 and 4, we split the overlapping sample by focusing on the two samples of deliveries from maternity units in the 250–749 and 750–1249 intervals, respectively.

The dichotomous variable 'Above thresholds', which equals one for units in the intervals 500–749 and 1000–1249 births, and zero for units in the intervals 250–499 and 750–999 births is positive and significant: being above the thresholds increases the probability of having a full team by about 29.5 percentage points in the full sample. Results from columns 3 and 4 confirm that being above the thresholds significantly increases the probability of a full team by 22.6 and 18 percentage points for the 500 and 1000 births cut-offs, respectively, in the full sample. Results are similar for the vaginal subsample, while in the subsample of C-sections the 500–749 dummy variable is not significant.

5.3. Delivery method

Fig. 3 shows the proportions of C-sections and emergency C-sections in the total number of births at the hospital-year level around the cut-off points. There seems to be no discrete jumps at any of the cut-offs, indicating that the mode of delivery is likely to be evenly distributed around the 500 and 1000 thresholds.

Tables 3 and 4 present the estimation results for Eqs. (1) and (3) respectively. Table 3 presents the results from a 2SLS estimation strategy, while Table A6 in the appendix shows the corresponding OLS estimates. Table 4 presents the RDD reduced form estimates.

In column (1) of Table 3, the dependent variable is C-section delivery, a binary variable equal to one if the birth was a caesarean and zero if it was a vaginal birth. The FT regressor for full team presence during delivery is positive but insignificant. In the following columns, we redefine the dependent variable using, alternatively, emergency C-section (column 2), planned C-section (column 3), assisted vaginal delivery (column 4) and uncomplicated vaginal delivery (column 5). We find that the presence of a full team is significant only in column 2: the presence of a full team reduces the probability of having an emergency C-section rather than a planned C-section or a vaginal delivery by about 5.9 percentage points, around one fourth of a standard deviation.

In Table 4 we repeat the analysis using the reduced form RDD. When the dependent variable is C-section, none of the cut-offs are significant in column 1. We find a significant negative sign for the overlapped discontinuity sample and above the 500 cut-off in column 2 for emergency C-sections. In column 4, when the dependent variable is assisted vaginal delivery, the coefficients are positive and significant above the cut-off in the overlapped sample and above 500 annual births. Taking the latter two results together, we find that the likelihood of an assisted vaginal delivery increases slightly, while the likelihood of an emergency C-section decreases more significantly in larger hospitals compared to any other type of delivery. The FT variable is not significant when mode of delivery is planned C-section (column 3) or uncomplicated vaginal

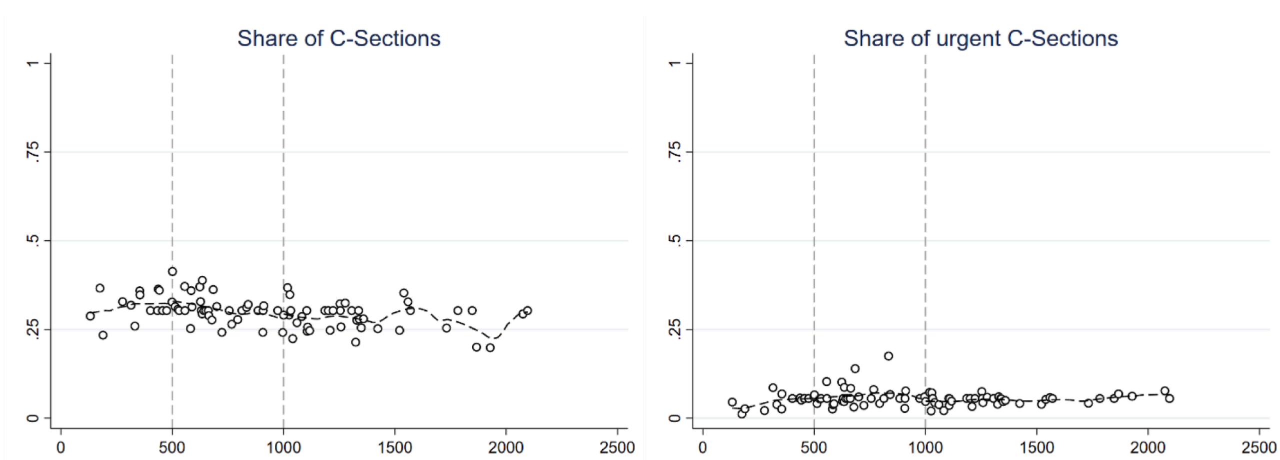

Fig. 3. Share of C-sections across thresholds.
Notes: The figure represents the average share of C-sections (left) and emergency C-sections (right) in the total number of births at the hospital-year level. The x-axis measures the number of births per year. Each point represents one hospital-year. The two thresholds are 500 and 1000 annual births. The total number of observations used in all panels is 55,840. The variables are defined in Online Appendix, Table A3.

delivery (column 5). One possible explanation is that larger hospitals are better able to cope with cases that require more staff effort. The choice between the two procedures depends crucially on the clinical scenario, the newborn's condition, the progress of labour and the mother's health. Emergency C-sections are likely to start as vaginal births and may be necessary in critical situations where the baby or the mother are at risk, or when vaginal labor is not progressing safely. However, when appropriate, an assisted vaginal delivery can avoid the risks of surgery and allow the mother to recover more quickly: it is less invasive and less costly, and it also reduces the risk of complications in future pregnancies.

5.4. Neonatal health outcomes

Fig. 4 plots neonatal outcomes against the annual number of births at the hospital level. If there is an effect, we would expect to see a discontinuous jump in outcomes at the cut-off points. Fig. 4 documents some discontinuities in the probability of some outcomes as a function of the annual number of births. For neonatal outcomes, there are discontinuous changes in the *No need for resuscitation* variable, in the Apgar scores (particularly at one minute), and the difference in Apgar scores at the cut-off points.

Tables 5 and 6 report the 2SLS and the RDD reduced form estimates for Eqs. (1) and (3), respectively. The dependent variable is one of the newborn health outcomes.[13]

In Table 5 we replicate the analysis on five different samples: the full sample of all deliveries (Panel A), the subsample of vaginal deliveries (Panel B), the subsample of uncomplicated vaginal deliveries (Panel C), the subsample of emergency C-sections (Panel D), and the subsample of planned C-sections (Panel E). Table A7 in the Online Appendix shows the corresponding OLS estimates for Eq. (1). For the full sample of all deliveries (panel A) we find that the presence of a full team (FT) increases the probability of no need for resuscitation (column 1) by almost 2.3 percentage points (around one-fifth of a standard deviation), while all Apgar scores (measured after one minute and after five minutes from birth, in columns 2 and 3) are lower (by 0.36 and 0.12 points, respectively) due to the presence of a full team. Similarly, the probability of having a perfect Apgar score (columns 4 and 5) decreases by about 8 and 4.5 percentage points, respectively. These results may be interpreted as evidence of more caution of full teams compared to smaller teams in attributing high Apgar scores, especially one minute after birth. In the medical literature, there is some evidence of interobserver variability of Apgar scores, across different birth settings and providers (O'Donnell et al., 2006). Some studies find a risk of bias of high Apgar scoring in the absence of independent checks against observer bias (Grünebaum et al., 2015; Wiegerinck et al., 2020).[14]

We consider two additional newborn health outcomes to explore potential mechanisms behind the effects observed on the two Apgar scores. Ideally, we would like to pinpoint the gynecologic, obstetric, or pediatric practices that a full team of professionals can perform better than a smaller team, which drives our results. First, we examine the difference between the Apgar score measured at five minutes and the Apgar score at one minute from birth (ΔApgar). We find that the difference between the five-minutes and the one-minute Apgar scores

Table 3
Effect of full team on delivery method – 2SLS estimates.

Dependent Variable:	*C-section delivery* (1)	*Emergency C-section* (2)	*Planned C-section* (3)	*Assisted vaginal delivery* (4)	*Uncomplicated vaginal Delivery* (5)
FT	0.0371	−0.0591***	0.0962	0.0085	−0.0456
	(0.084)	(0.018)	(0.082)	(0.013)	(0.085)
Mean of Y	0.289	0.0512	0.238	0.0404	0.670
SD of Y	0.453	0.220	0.426	0.197	0.470
F-Stat	35.63	35.63	35.63	35.63	35.63
J-Stat	0.0786	3.005	0.00736	7.399	0.517
J-pval	0.779	0.0830	0.932	0.00652	0.472
Observations	55,840	55,840	55,840	55,840	55,840

Notes: Each column presents the IV estimation of Eq. (1). The dependent variables are: C-section delivery (equal to one if the delivery was by a C-Section, and zero for a vaginal delivery) in column (1); Emergency C-section (equal to one if the delivery was an emergency unscheduled C-section, and zero for a planned C-section delivery or a vaginal delivery) in column (2); Planned C-section (equal to one if the delivery was a planned scheduled C-section, and zero for an emergency C-section or a vaginal delivery) in column (3); Assisted vaginal delivery (equal to one if the delivery was vaginal and performed with the help of forceps or a vacuum device, and zero for a C-section delivery or an unassisted vaginal delivery) in column (4); Uncomplicated vaginal delivery (equal to one if the delivery was vaginal and spontaneous, and zero for a C-section delivery or an assisted vaginal delivery) in column (5). FT is the Full team variable. In all specifications, we also include mothers' characteristics (age, nationality, education level, employment status, marital status of the mother, whether the woman is at her first delivery, delivery week, whether the woman experienced hospitalizations during the pregnancy, the weight gain during the pregnancy, smoking, experiences of previous abortions and previous miscarriage, distance to the closest hospital), delivery characteristics (newborn head circumference, monitoring of fetal heartbeat, weekend delivery, night shift delivery, antibiotics during labor), hospital characteristics (the ratio of femoral neck fractures treated within two days), local health authority, month, and year fixed effects. The instrumental variables are dummies for the two thresholds '500–999′ and 'Above 1000'. The standard errors are clustered at the hospital, shift, and weekdays level. *Mean of Y* and *SD of Y* report the dependent variable's sample mean and standard deviation, respectively. *F-stat* is the F-statistics on the external instrumental variables in the first stage regression, while *J-stat* and *J-pval* are the Sargan-Hansen J statistics, and the corresponding p-value. *Observations* report the number of observations included in the estimation. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(column 6) increases by 0.23 points with a full team. Our explanation is that the presence of a full team can significantly improve a newborn's health in the few minutes after birth compared to a smaller team. In particular, the presence of a pediatrician assisting at birth may have a positive effect on health status because of their expertise. In addition, the presence of the pediatrician allows the other professionals in the team (gynecologists and midwives) to concentrate more on their specific tasks. Second, we consider the presence of meconium, which may cause fetal distress if the newborn aspirates it during labor or birth. The hypothesis that a full team may be more capable of preventing such an adverse event than a smaller team is not supported by our estimates, as the probability of no meconium (column 7) does not significantly change with a full team.

We repeat the analysis on the sub-samples of vaginal (panels B and C) and C-section deliveries (panels D and E). We find that results in the full sample are mainly driven by the results in the vaginal sub-samples, while there are no significant effects in the C-sections sub-samples. For vaginal deliveries in panel B, we find that with a full team, there is a higher probability of no need for resuscitation (+1.7 percentage points, around one-fifth of a standard deviation), lower Apgar scores (at one and five minutes after birth), and a higher difference between the five-minute and one-minute Apgar scores (+0.26 points, almost one-third of a standard deviation). Results are similar for the uncomplicated vaginal delivery subsamples of panel C.

Overall, the full team variable is never significant in the sub-samples

[13] The sample size varies due to incomplete data on neonatal health outcomes. We therefore conducted a sensitivity analysis using the same sample size for all neonatal health outcomes (see Tables 5 and 6). The estimates are available upon request and the main results remain consistent when the sample size is kept the same.

[14] Tables A7 in the appendix shows the OLS results. All coefficients are statistically significant in almost every subsample. In addition, while the coefficients for no need for resuscitation and no meconium are negative, the coefficients for all Apgar scores are quite similar in size and magnitude to the 2SLS results.

Table 4
Effect of full team on delivery method – RDD reduced form estimates.

Dependent Variable:	*C-section delivery* (1)	*Emergency C-section* (2)	*Planned C-section* (3)	*Assisted vaginal delivery* (4)	*Uncomplicated vaginal delivery* (5)
Panel A. Discontinuity sample					
Above threshold	0.0048	−0.0182**	0.0230	0.0107*	−0.0155
	(0.025)	(0.008)	(0.026)	(0.006)	(0.024)
Mean of Y	0.295	0.0514	0.243	0.0413	0.664
SD of Y	0.456	0.221	0.429	0.199	0.472
Observations	35,067	35,067	35,067	35,067	35,067
Panel B. Sample 250–749 yearly births					
Above threshold	−0.0245	−0.0401***	0.0156	0.0206**	0.0039
	(0.039)	(0.009)	(0.041)	(0.009)	(0.036)
Mean of Y	0.322	0.0531	0.269	0.0385	0.640
SD of Y	0.467	0.224	0.443	0.192	0.480
Observations	14,281	14,281	14,281	14,281	14,281
Panel C Sample 750–1249 yearly births					
Above threshold	0.0233	−0.0021	0.0255	0.0069	−0.0303
	(0.041)	(0.011)	(0.043)	(0.008)	(0.041)
Mean of Y	0.278	0.0503	0.228	0.0430	0.679
SD of Y	0.448	0.219	0.420	0.203	0.467
Observations	20,786	20,786	20,786	20,786	20,786

Notes: Each column presents the reduced form estimates for Eq. (3). Panel A shows results from the sample, including all deliveries in maternity units that record yearly births within the ±250 bandwidth around the thresholds of 500 and 1000. In Panel B, we include all deliveries in maternity units that record yearly births within the ±250 bandwidth around the 500 threshold (250–749). In Panel C, we include all deliveries in maternity units that record yearly births within the ±250 bandwidth around the 1000 threshold (750–1250). In all specifications, we include the following set of controls: the difference (and its square) between the total number of births at the hospital-year level and the closest threshold (500 or 1000); mothers' characteristics (age, nationality, education level, employment status, marital status of the mother, whether the woman is at her first delivery, delivery week, whether the woman experienced hospitalizations during the pregnancy, the weight gain during the pregnancy, smoking, experiences of previous abortions and previous miscarriage, distance to the closest hospital), delivery characteristics (newborn head circumference, monitoring of fetal heartbeat, weekend delivery, night shift delivery, antibiotics during labor), hospital characteristics (the ratio of femoral neck fractures treated within two days), local health authority, month, and year fixed effects. The standard errors are clustered at the hospital, shift, and weekdays level. *Mean of Y* and *SD of Y* report the dependent variable's sample mean and standard deviation, respectively. *Observations* report the number of observations included in the estimation. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

of emergency and planned C-sections. Our preferred explanation is that the production process for these two procedures is likely to be different due to the highly specialized surgical team, the specific surgical environment and instruments required for a C-section, and generally the different levels of preparation and resources required for the two modes of delivery. The organizational and staffing requirements for vaginal and C-section deliveries differ in terms of the staff and equipment needed to ensure the safety and well-being of both mother and newborn. While a vaginal delivery usually takes place in a labor room equipped with standard obstetric tools and monitoring systems, a C-section takes place in an operating room with strict sterile conditions and additional sterilized equipment, anesthesia machines, and specialized surgical instruments. In addition, because of its surgical nature, a C-section requires a larger, more specialized team, typically an obstetric surgeon, an anesthetist, surgical assistants, and operating room nurses. While our data allow us to identify the role of a full team for vaginal deliveries, we are unlikely to be able to fully consider all the specifics of C-sections. For this reason, we will mainly focus on the uncomplicated vaginal delivery subsample in the RDD specification.

Table 6 presents results for the reduced form RDD specification of Eq. (3) for the subsample of uncomplicated vaginal births.[15] The observations we consider are further restricted to maternity units whose number of births is close to the discontinuity points (Angrist and Lavy, 1999). We define three "discontinuity samples" that include only maternity units whose number of births is in the interval ±250 around the cut-offs 500 and 1000, and an overlapped sample. This allows us to draw several insights from a set of maternity wards that are very similar in workload, but differ in staff units.[16] The variable of interest is a dichotomous variable for being above the threshold (500–749 births or 1000–1249 births per year) in Panel A of Table 6; while it is a dichotomous variable equal to one if the number of yearly births is in the 500–749 interval in Panel B, and finally, a dummy variable equal to one if the number of births is in the interval 1000–1249 births, in Panels C.

As can be seen in Table 6, most of the dichotomous variables for the thresholds have consistent signs across the different samples, although the magnitudes and significance levels differ. Furthermore, the results are consistent with those in Table 5, where we use larger samples and a 2SLS strategy. We find that being above a threshold increases the likelihood of not needing resuscitation relative to the comparable group of maternity units just below the same threshold, but only significantly so above the 500 threshold. We also find that Apgar scores (at one minute and five minutes) and the probability of a perfect Apgar score (greater than nine) are lower for births above both thresholds. However, the effect is always significant only above the 1000 threshold. Being above the cut-off improves the difference between the 5-minute Apgar and the 1-minute Apgar for each cut-off. Finally, the probability of no meconium is lower above the 500 cut-off and higher above the 1000 cut-off.

5.5. Maternal health outcomes

Fig. 5 plots maternal health outcomes against the annual number of births at the hospital level. We find some discontinuous jumps for lacerations and episiotomy.

Tables 7 and 8 present the 2SLS and RDD estimates for Eqs. (1) and (3), when we consider the maternal health status as a dependent variable.

For the full sample of deliveries (Panel A of Table 7), we find a positive effect for a full team on the probability of not experiencing a second/third-degree laceration (column 2) and a negative effect for the *No episiotomy* outcome (column 3). More precisely, women assisted by a full team have a higher probability of undergoing an episiotomy by about 17.4 percentage points. However, more severe lacerations (second or third-degree) are reduced with a full team: the probability of having

[15] The full estimation results are presented in Appendix Table A11. Results for the other subsamples (full sample, all vaginal deliveries, emergency C-sections and planned C-sections) are presented in Appendix Table A9.

[16] To a lesser extent they also differ in the availability of equipment and other facilities. See the discussion in Section 5.1.

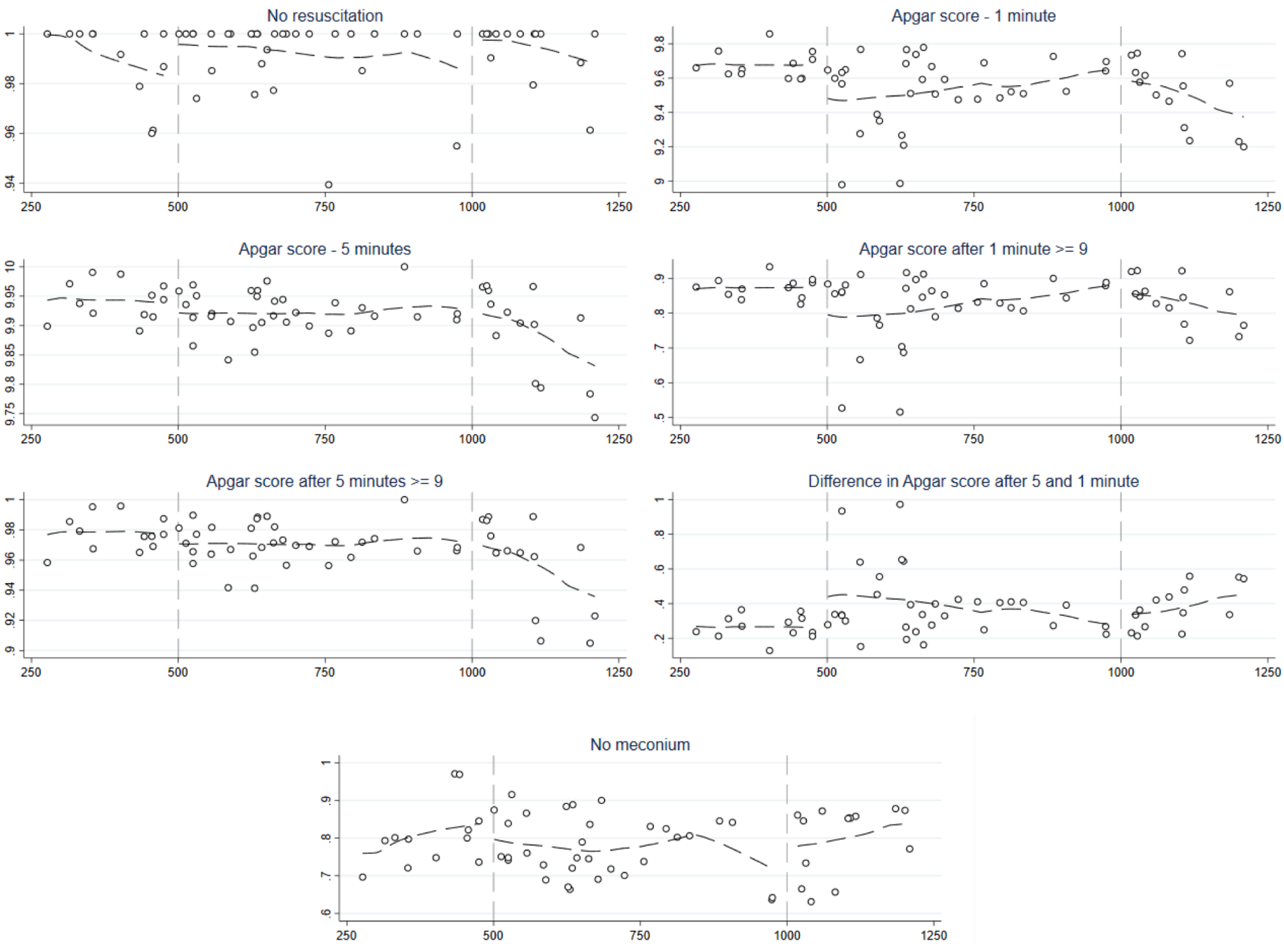

Fig. 4. Neonatal health outcomes across thresholds.
Notes: Each panel represents the average level of newborn health outcomes (for the sample of vaginal births) at hospital level. The neonatal health outcomes considered are: (i) no need for resuscitation (number of observations 24,818); (ii) Apgar score at 1 min (number of observations 24,029); (iii) Apgar score at 5 min (number of observations 24,028); (iv) proportion of Apgar scores at 1 min greater than 9 (number of observations 24,029); (v) proportion of Apgar scores at 5 min greater than 9 (number of observations 24,028); (vi) difference between Apgar scores at 5 min and 1 min (number of observations 24,017); (vii) proportion of births without meconium (number of observations 24,818). The y-axis measures health outcomes. The x-axis measures the number of births per year. Each point represents a hospital-year. The two thresholds are at 500 and 1000 annual births. Variables are defined in the Online Appendix, Table A3.

no second or third-degree lacerations increases by about 9.2 percentage points with a full team. All in all, severe lacerations are avoided thanks to a more intensive surgical intervention that seems to be appropriate in this case. We find no effects of a full team on other complications measured by a longer hospital stay (column 4). These findings for maternal health outcomes are confirmed for the vaginal birth subsamples (Panels B and C of Table 7), but not for the C-section subsamples (where applicable, Panels D and E of Table 7), where the full team variable is never significant.

Table 8 repeats the same analysis for maternal health outcomes in the uncomplicated vaginal births subsample, using reduced RDD estimates of Eq. (3).[17] We find a positive effect on the probability of having no severe lacerations (column 2), which is statistically significant above the 1000 cut-off. We also find more episiotomies, significant in the overlapped sample and above the 500 cut-off. Finally, the probability of no complications increases above the threshold of 1000 births.

Overall, we find that vaginal births above the cut-off are associated with better neonatal and maternal health outcomes in terms of less need for resuscitation, lower Apgar scores that improve between 1 and 5 min, and fewer obstetric lacerations, which is associated with more appropriate episiotomy. The results are consistent for the overlapped sample and the two samples around the 500 and 1000 thresholds, although the magnitude and significance levels may differ. All else being equal, the coefficients are usually larger for larger hospitals around the 1000 threshold, confirming the positive association between volume and outcomes, as larger hospitals ensure greater productivity, as measured by better health outcomes (e.g., Gaynor et al., 2005; Mesman et al., 2015; Avdic et al. 2024). We also find that there are other positive outcomes above the 1000 threshold that are not present above the 500 threshold: namely lower incidence of meconium and shorter length of hospital stay for both mothers and newborns.[18]

[17] The full estimation results are presented in Online Appendix Table A11. Results for the other subsamples (full sample, all vaginal deliveries, emergency C-sections and planned C-sections) are presented in Online Appendix Table A10.

[18] Moreover, if we couple the lower Apgar scores when a full team is present with the fact that larger hospitals above the 1,000 threshold treat worse cases (as discussed in Section 5.1 and Table 1), we can infer that the identified effect represents a lower bound on the true positive effect of a full team on health outcomes, since it includes the negative effect on health of the worse patient population treated in hospitals with a full team. We are grateful to a reviewer for pointing this out.

Table 5
Effect of full team on neonatal health outcomes - 2SLS estimates.

	No resusc (1)	Apgar1 (2)	Apgar5 (3)	Apgar1≥9 (4)	Apgar5≥9 (5)	ΔApgar (6)	NoMecon (7)
Panel A - All deliveries							
FT	0.0227***	−0.3605***	−0.1233**	−0.0821**	−0.0447**	0.2295***	−0.0181
	(0.007)	(0.121)	(0.051)	(0.038)	(0.018)	(0.083)	(0.042)
Mean of Y	0.991	9.527	9.885	0.837	0.958	0.359	0.842
SD of Y	0.0969	1.227	0.623	0.370	0.200	0.959	0.365
F-Stat	35.42	35.96	35.97	35.96	35.97	35.93	35.42
J-Stat	4.049	0.659	5.688	0.755	6.203	0.0328	5.810
J-pval	0.0442	0.417	0.0171	0.385	0.0128	0.856	0.0159
Observations	55,840	54,218	54,209	54,218	54,209	54,181	55,840
Panel B - Vaginal deliveries							
FT	0.0173**	−0.4030***	−0.1369***	−0.1110***	−0.0512***	0.2628***	−0.0179
	(0.007)	(0.130)	(0.049)	(0.043)	(0.019)	(0.091)	(0.050)
Mean of Y	0.992	9.582	9.902	0.853	0.964	0.320	0.807
SD of Y	0.0906	1.142	0.566	0.355	0.187	0.906	0.395
F-Stat	20.87	21.53	21.55	21.53	21.55	21.52	20.87
J-Stat	2.529	0.00165	1.889	0.0326	1.895	0.558	6.974
J-pval	0.112	0.968	0.169	0.857	0.169	0.455	0.00827
Observations	39,839	38,763	38,753	38,763	38,753	38,738	39,839
Panel C -Uncomplicated Vaginal deliveries							
FT	0.0141**	−0.3476***	−0.1154***	−0.1007***	−0.0407***	0.2281***	−0.0021
	(0.006)	(0.108)	(0.040)	(0.038)	(0.015)	(0.080)	(0.050)
Mean of Y	0.993	9.634	9.914	0.868	0.969	0.280	0.812
SD of Y	0.0845	1.064	0.537	0.338	0.174	0.854	0.390
F-Stat	18.77	19.39	19.40	19.39	19.40	19.37	18.77
J-Stat	1.754	0.00171	2.324	0.152	2.072	0.454	7.126
J-pval	0.185	0.967	0.127	0.697	0.150	0.500	0.00760
Observations	37,726	36,727	36,718	36,727	36,718	36,704	37,726
Panel D - Emergency C-Sections							
FT	0.0527	−0.4969	−0.0252	−0.0467	0.0271	0.4257	0.0667
	(0.050)	(0.553)	(0.355)	(0.137)	(0.112)	(0.381)	(0.110)
Mean of Y	0.978	9.033	9.701	0.708	0.900	0.681	0.870
SD of Y	0.148	1.800	1.018	0.455	0.301	1.274	0.336
F-Stat	14.53	13.85	13.77	13.85	13.77	13.86	14.53
J-Stat	3.914	0.00118	2.629	0.207	2.301	1.243	2.019
J-pval	0.0479	0.973	0.105	0.649	0.129	0.265	0.155
Observations	2815	2718	2722	2718	2722	2716	2815
Panel E - Planned C-Sections							
FT	0.0240	−0.1987	−0.0884	0.0028	−0.0275	0.0792	−0.0509
	(0.016)	(0.207)	(0.091)	(0.071)	(0.028)	(0.165)	(0.034)
Mean of Y	0.990	9.471	9.875	0.818	0.956	0.405	0.941
SD of Y	0.101	1.289	0.660	0.386	0.206	1.018	0.235
F-Stat	14.64	14.83	14.80	14.83	14.80	14.84	14.64
J-Stat	3.282	5.020	10.76	2.033	13.01	1.592	0.0362
J-pval	0.0700	0.0251	0.00103	0.154	0.000310	0.207	0.849
Observations	13,186	12,708	12,705	12,708	12,705	12,698	13,186

Notes: Each column presents the IV estimation of Eq. (1). In Panel A we consider the sample of all deliveries, in Panel B we restrict the sample to vaginal deliveries only, in panels C, D, and E the samples consist of Uncomplicated vaginal deliveries, Emergency C-sections and Planned C-sections, respectively. The dependent variables are: No Resusc (equal to one if the newborn did not need any resuscitation, and zero otherwise) in column (1); Apgar1 for Apgar scores after one minute (ranging from 0 to 10) in column (2); Apgar5 for Apgar scores after five minutes (ranging from 0 to 10) in column (3); Apgar1 ≥9 (equal to one for Apgar 1 greater than nine, and zero otherwise), in column (4); and Apgar5≥9 (equal to one for Apgar 5 greater than nine, and zero otherwise), in column(5); ΔApgar (the difference between Apgar5 and Apgar1) in column (6); No Mecon (a binary variable equal to one if no meconium is present at birth, and zero otherwise) in column (7). FT is the Full team variable. In all specifications, we also include mothers' characteristics (age, nationality, education level, employment status, marital status of the mother, whether the woman is at her first delivery, delivery week, whether the woman experienced hospitalizations during the pregnancy, the weight gain during the pregnancy, smoking, experiences of previous abortions and previous miscarriage, distance to the closest hospital), delivery characteristics (newborn head circumference, monitoring of fetal heartbeat, weekend delivery, night shift delivery, antibiotics during labor), hospital characteristics (the ratio of femoral neck fractures treated within two days), local health authority, month, and year fixed effects. The instrumental variables are dummies for the two thresholds '500–999′ and 'Above 1000'. The standard errors are clustered at the hospital, shift, and weekdays level. *Mean of Y* and *SD of Y* report the dependent variable's sample mean and standard deviation, respectively. *F-stat* is the F-statistics on the external instrumental variables in the first stage regression, while *J-stat* and *J-pval* are the Sargan-Hansen J statistics, and the corresponding p-value. *Observations* report the number of observations included in the estimation. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.6. Sensitivity analysis and heterogeneous effects

As the results hold primarily for uncomplicated vaginal deliveries, we focus on these births and report the results for some additional outcomes in Table 9. First, breastfeeding (column 1) decreases above the thresholds. This is significant for the overlapped sample and above the 500 threshold. The probability of not using a Kristeller maneuver (column 2) increases in the discontinuity sample and above the 500 birth threshold, and it decreases above the 1000 cutoff. The probability of spontaneous afterbirth (column 3) decreases above the 500 threshold, but it increases above the 1000 cutoff. We also find that the probability of not using oxytocin (column 4), prostaglandins (column 5), and amniorrhexis (i.e., the rupture of membranes to facilitate delivery, column 6) decreases at all thresholds.

These results indicate that larger hospitals make more use of treatments to facilitate birth than smaller hospitals. In other words, larger hospitals "medicalize" deliveries more than smaller ones, managing and controlling the natural birth process with medical interventions, such as the use of oxytocin and prostaglandins to stimulate labor or induce rupture of the membranes, which in turn may hinder breastfeeding. As

Table 6
Effect of full team on neonatal health outcomes – RDD reduced form estimates on the uncomplicated vaginal deliveries sample.

	No resusc (1)	Apgar1 (2)	Apgar5 (3)	Apgar1≥9 (4)	Apgar5≥9 (5)	ΔApgar (6)	NoMecon (7)
Panel A - Discontinuity sample							
Above threshold	0.0023	−0.0783**	−0.0252	−0.0205	−0.0107*	0.0494*	0.0024
	(0.003)	(0.038)	(0.016)	(0.015)	(0.006)	(0.030)	(0.024)
Mean of Y	0.993	9.612	9.920	0.860	0.971	0.309	0.789
SD of Y	0.0849	1.088	0.515	0.347	0.169	0.892	0.408
Observations	23,435	22,692	22,691	22,692	22,691	22,680	23,435
Panel B - Sample 250–749 births							
Above threshold	0.0078*	−0.0459	0.0001	−0.0059	−0.0040	0.0361	−0.0793**
	(0.004)	(0.048)	(0.022)	(0.014)	(0.007)	(0.034)	(0.037)
Mean of Y	0.994	9.651	9.939	0.866	0.978	0.289	0.777
SD of Y	0.0751	0.987	0.459	0.341	0.147	0.837	0.416
Observations	9194	9007	9009	9007	9009	9004	9194
Panel C - Sample 750–1249 births							
Above threshold	0.0021	−0.2035***	−0.0583*	−0.0962***	−0.0284*	0.1459**	0.0771*
	(0.004)	(0.067)	(0.032)	(0.026)	(0.015)	(0.057)	(0.043)
Mean of Y	0.992	9.588	9.909	0.856	0.966	0.320	0.796
SD of Y	0.0901	1.142	0.545	0.351	0.181	0.923	0.403
Observations	14,241	13,685	13,682	13,685	13,682	13,676	14,241

Notes: Each column presents the reduced form estimates for Eq. (3). Panel A shows results from the sample, including all uncomplicated vaginal deliveries in maternity units that record yearly births within the ±250 bandwidth around the thresholds of 500 and 1000. In Panel B, we include all uncomplicated vaginal deliveries in maternity units that record yearly births within the ±250 bandwidth around the 500 threshold (250–749). In Panel C, we include all uncomplicated vaginal deliveries in maternity units that record yearly births within the ±250 bandwidth around the 1000 threshold (750–1250). In all specifications, we include the following set of controls: the difference (and its square) between the total number of births at the hospital-year level and the closest threshold (500 or 1000); mothers' characteristics (age, nationality, education level, employment status, marital status of the mother, whether the woman is at her first delivery, delivery week, whether the woman experienced hospitalizations during the pregnancy, the weight gain during the pregnancy, smoking, experiences of previous abortions and previous miscarriage, distance to the closest hospital), delivery characteristics (newborn head circumference, monitoring of fetal heartbeat, weekend delivery, night shift delivery, antibiotics during labor), hospital characteristics (the ratio of femoral neck fractures treated within two days), local health authority, month, and year fixed effects. The standard errors are clustered at the hospital, shift, and weekdays level. *Mean of Y* and *SD of Y* report the dependent variable's sample mean and standard deviation, respectively. *Observations* report the number of observations included in the estimation. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

before, we interpret the "medicalization" of labor in the direction of reducing the risk of complications for both mother and newborn and increasing safety by having medical options available in case of emergencies.

Table A12 in the Online Appendix shows a placebo analysis considering other neonatal health outcomes that should not be affected by the presence of a full team. The outcomes considered are: 1) the probability of low birth weight, i.e. a binary variable equal to one for birth weight <2500 g (in column 1); 2) birth weight in grams (in column 2); and 3) gestational age in weeks (in column 3). The impact of the thresholds on birth weight is positive and significant in panels A and C: birth weight is about 39–48 g (about one-tenth of the sample standard deviation) higher above the thresholds. None of the other outcomes vary discontinuously around the threshold, confirming that in general, only health outcomes that should be influenced by increased health worker productivity are significantly affected.

Finally, we examine the presence of heterogeneous effects by looking at births that occur at different times of the day and week and by taking into account the nationality of the mothers. First, we look at the descriptive statistics for the presence of a full team during weekends and weekdays, night shifts and day shifts (last four rows of Online Appendix Table A4, Panel B). We find that the presence of a full team is less likely on weekends and night shifts than on weekdays and day shifts, usually assisting in around less than a third of cases around the 500 cut-off and half of cases around the 1000 cut-off.

Our RDD estimation results are presented in the Online Appendix Table A13 for neonatal outcomes and Table A14 for maternal outcomes. We split our sample of vaginal births to compare the following: 1) births during weekends and holidays (Panel A of Tables A13 and A14) as opposed to working days (Panel B of Tables A13 and A14), and 2) the day shift from 8am to 11.59pm. (Panel C of Tables A13 and A14) as opposed to the night shift from midnight to 7.59am (Panel D of Tables A13 and A14). We also consider the two subsamples of native and non-native mothers (Panels E and F of Tables A13 and A14). The reduced number of observations in the sub-samples reduces the precision of many estimates. Considering neonatal outcomes first, we do not find clear heterogeneous patterns between weekends and workdays, between day and night shifts, nor between native and non-native mothers. The coefficient of the 'above threshold' binary variable, when the dependent variable is 'no need for resuscitation' (column 1 of Table A13), is almost always positive. However, it is only significant above the thresholds during the day shift and for samples of foreign-born women. The coefficients of the 'above threshold' variable when the dependent variable is Apgar scores (columns 2–5) are mostly negative and significant primarily above the threshold of 1000. When the dependent variable is the difference in Apgar scores, the coefficient of the 'above threshold' variable is always positive (column 6) and is estimated with precision for workdays, day shifts and native mothers, particularly above the 1000 cut-off. Finally, meconium (column 7) is more likely to occur above the 500 cut-off, but less likely above the 1000 cut-off. Overall, the coefficients when considering the impact of the thresholds are very similar for weekends and working days or day and night shifts. The differences, if any, are mainly related to the threshold considered, with better results above the 1000 threshold compared to the 500 threshold in almost all subsamples. The results are also consistent across the sub-samples of native and non-native mothers, with greater precision in the estimates for the larger sample of native mothers.

Table A14 presents the same analysis for maternal health outcomes. Also in this case, we do not find any substantial heterogeneity. Most of the coefficients when estimating models for lacerations (columns 1 and 2) are not precisely estimated. Episiotomy (column 3) is more likely (and significantly so) above the 500 cut-off, especially during day shifts as opposed to night shifts and for non-native mothers (in terms of magnitude) as opposed to natives. Finally, the absence of complications (i.e., longer hospital stay after delivery) is less likely above the 500 cut-off for births during day shifts and for native mothers.

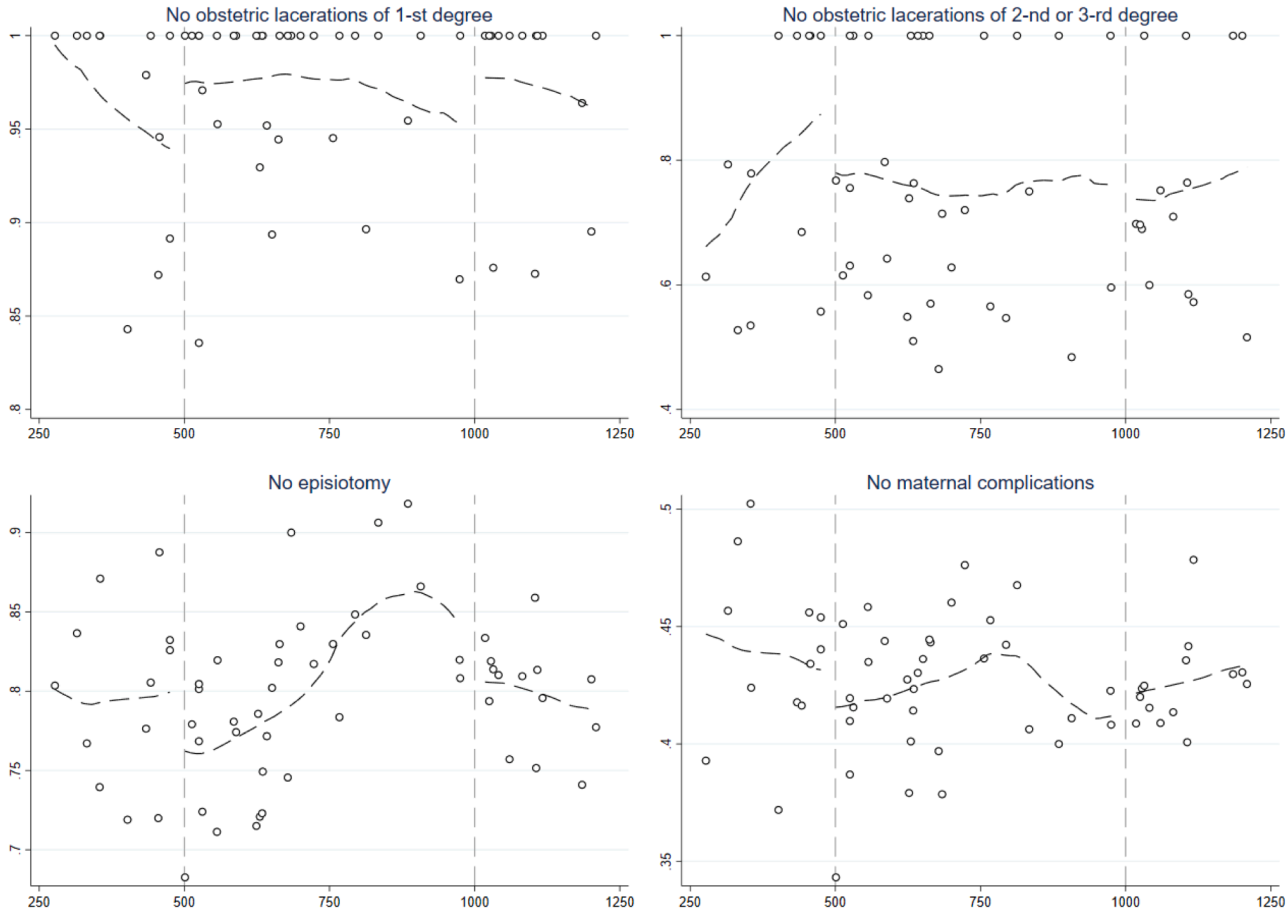

Fig. 5. Maternal health outcomes across thresholds.

Notes: Each panel represents the average level of maternal health outcomes (for the sample of vaginal births) at hospital level. The maternal health outcomes considered are: (i) no obstetric laceration of 1st degree; (ii) no obstetric laceration of 2nd or 3rd degree; (iii) no episiotomy; (iv) no maternal complications. The y-axis measures health outcomes. The x-axis measures the number of births per year. Each point represents a hospital-year. The two thresholds are at 500 and 1000 annual births. The total number of observations used in all panels is 24,818. Variables are defined in the Online Appendix, Table A3.

6. Conclusions

With the expected growing demand for healthcare services and additional pressures to spending growth from rapid technological innovation, better understanding the role of medical staff and its effects on productivity (measured by patient health outcomes) has become a priority for policymakers worldwide. A growing literature in economics is exploring medical staff productivity along several dimensions. In this paper, we exploit an Italian regulation that allows us to avoid the endogeneity problem around medical labor to estimate the productivity of a full team of professionals in maternity wards in terms of the choice of delivery method and the general health status of newborns and mothers.

First, we show that a full team of professionals (a midwife, a gynecologist and a pediatrician) is more likely to be present at delivery in larger hospitals. We find no role for the full team in the choice of mode of delivery. Finally, we find that a full team is associated with better health outcomes. In particular, we detect that a full team is associated with more medical treatments used to prevent complications during labour and delivery for both mother and newborn.

Our findings carry important policy implications. First, as hospitals above the cut-offs imposed by regulations relating to the size of the maternity unit obtain better health outcomes (hence, they are more productive) than those below, we provide additional support to the view that policymakers should concentrate hospital activities in larger and more specialized units. This observation calls to carefully scrutinize and, possibly, close maternity wards below the 500-childbirths-per-year threshold. However, as the closures are likely to affect patients in remote areas (e.g., Perucca et al., 2019), who would have to travel further to seek maternity care, there is the need to define appropriate mechanisms to avoid closures that increase disparities of opportunity and, more importantly, risks for neonatal and maternal health even in the short run, before patients and hospitals can adjust to closures (e.g., Avdic et al., 2024; Avdic, 2016).

Second, our findings suggest that one possible underlying mechanism of the volume-outcome relationship is linked to physician and hospital characteristics, such as the skills and availability of specific human resources (e.g., Mesman et al., 2015). In particular, the reduced availability of clinical staff in hospitals slightly below statutory thresholds can create an artificial difference relative to hospitals with very similar workloads but slightly above the cut-offs. This calls for more flexible rules instead of fixed thresholds in the definition of staff needs for each maternity ward.

Original publication

We declare no prior or duplicate publication or submission elsewhere of any part of the work included in the manuscript.

Ethics

We declare no ethical issues are involved in the production of this manuscript.

Table 7
Effect of full team on maternal health outcomes - 2SLS estimates.

	No obstetric lacerations 1st degree	No obstetric lacerations 2nd-3rd degree	No episiotomy	No other complications
	(1)	(2)	(3)	(4)
Panel A - All deliveries				
FT	0.0007	0.0919***	−0.1737**	−0.0194
	(0.010)	(0.031)	(0.071)	(0.023)
Mean of Y	0.976	0.826	0.560	0.437
SD of Y	0.155	0.379	0.496	0.496
F-Stat	35.42	35.42	35.42	35.42
J-Stat	0.0750	2.416	0.0159	0.000207
J-pval	0.784	0.120	0.900	0.989
Observations	55,840	55,840	55,840	55,840
Panel B - Vaginal deliveries				
FT	−0.0026	0.0929***	−0.2046***	−0.0143
	(0.012)	(0.031)	(0.066)	(0.027)
Mean of Y	0.966	0.755	0.787	0.433
SD of Y	0.182	0.430	0.410	0.496
F-Stat	20.87	20.87	20.87	20.87
J-Stat	0.757	3.090	0.0291	0.529
J-pval	0.384	0.0788	0.865	0.467
Observations	39,839	39,839	39,839	39,839
Panel C - Uncomplicated Vaginal deliveries				
FT	−0.0013	0.0827***	−0.2074***	−0.0170
	(0.012)	(0.030)	(0.068)	(0.028)
Mean of Y	0.964	0.742	0.775	0.434
SD of Y	0.187	0.437	0.418	0.496
F-Stat	18.77	18.77	18.77	18.77
J-Stat	0.344	3.411	0.0767	0.339
J-pval	0.558	0.0648	0.782	0.560
Observations	37,726	37,726	37,726	37,726
Panel D - Emergency C-Sections				
Full team	–	–	−0.0118	−0.0160
			(0.044)	(0.157)
Mean of Y	1	0.999	0.0137	0.430
SD of Y	0.0171	0.0382	0.116	0.495
F-Stat	–	–	14.53	14.53
J-Stat	–	–	0.959	0.596
J-pval			0.327	0.440
Observations	2815	2815	2815	2815
Panel E - Planned C-Sections				
FT	–	–	0.0032	−0.0540
			(0.002)	(0.058)
Mean of Y	1	1	0.000815	0.448
SD of Y	0.00792	0.0137	0.0285	0.497
F-Stat	–	–	14.64	14.64
J-Stat	–	–	0.782	3.704
J-pval			0.376	0.0543
Observations	13,186	13,186	13,186	13,186

Notes: Each column presents the IV estimation results for Eq. (1). In Panel A we consider the sample of all deliveries, in Panel B we restrict the sample to vaginal deliveries only, in panels C, D, and E the samples consist of Uncomplicated vaginal deliveries, Emergency C-sections and Planned C-sections, respectively. The dependent variables are No Obstetric lacerations - 1st degree (equal to one if the woman did not experience any 1st-degree lacerations, and zero otherwise) in column (1); No Obstetric lacerations - 2nd-3rd degree (equal to one if the woman did not experience any 2nd/3rd-degree lacerations, and zero otherwise) in column (2); No episiotomy (equal to one if the woman did not experience any episiotomy) in column (3); No other complications (equal to one if the hospital stay was equal to or less than two days, and zero otherwise), in column (4). FT is the Full Team variable. In all specifications, we also include mothers' characteristics (age, nationality, education level, employment status, marital status of the mother, whether the woman is at her first delivery, delivery week, whether the woman experienced hospitalizations during the pregnancy, the weight gain during the pregnancy, smoking, experiences of previous abortions and previous miscarriage, distance to the closest hospital), delivery characteristics (newborn head circumference, monitoring of fetal heartbeat, weekend delivery, night shift delivery, antibiotics during labor), hospital characteristics (the ratio of femoral neck fractures treated within two days), local health authority, month, and year fixed effects. The instrumental variables are dummies for the two thresholds '500–999′ and 'Above 1000'. The standard errors are clustered at the hospital, shift, and weekdays level. *Mean of Y* and *SD of Y* report the dependent variable's sample mean and standard deviation, respectively. ***F-stat*** is the F-statistics on the external instrumental variables in the first stage regression, while *J-stat* and *J-pval* are the Sargan-Hansen J statistics, and the corresponding p-value. *Observations* report the number of observations included in the estimation. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8
Effect of full team maternal health outcomes – RDD Reduced form estimates on the uncomplicated vaginal deliveries sample.

	No obstetric lacerations 1st degree	No obstetric lacerations 2nd-3rd degree	No episiotomy	No other complications
	(1)	(2)	(3)	(4)
Panel A Discontinuity sample				
Above threshold	0.0063	0.0282	−0.0469**	0.0036
	(0.007)	(0.021)	(0.018)	(0.014)
Mean of Y	0.965	0.753	0.788	0.428
SD of Y	0.183	0.431	0.409	0.495
Observations	23,435	23,435	23,435	23,435
Panel B Sample 250–749 births				
Above threshold	−0.0057	−0.0001	−0.1046***	−0.0181
	(0.009)	(0.031)	(0.020)	(0.020)
Mean of Y	0.973	0.760	0.775	0.429
SD of Y	0.163	0.427	0.417	0.495
Observations	9194	9194	9194	9194
Panel C Sample 750–1249 births				
Above threshold	−0.0096	0.0747**	−0.0031	0.0527**
	(0.013)	(0.034)	(0.020)	(0.023)
Mean of Y	0.961	0.749	0.795	0.427
SD of Y	0.194	0.434	0.404	0.495
Observations	14,241	14,241	14,241	14,241

Notes: Each column presents the reduced form estimates for Eq. (3). Panel A shows results from the sample, including all uncomplicated vaginal deliveries in maternity units that record yearly births within the ±250 bandwidth around the thresholds of 500 and 1000. In Panel B, we include all uncomplicated vaginal deliveries in maternity units that record yearly births within the ±250 bandwidth around the 500 threshold (250–749). In Panel C, we include all uncomplicated vaginal deliveries in maternity units that record yearly births within the ±250 bandwidth around the 1000 threshold (750–1249). In all specifications, we include the following set of controls: the difference (and its square) between the total number of births at the hospital-year level and the closest threshold (500 or 1000); mothers' characteristics (age, nationality, education level, employment status, marital status of the mother, whether the woman is at her first delivery, delivery week, whether the woman experienced hospitalizations during the pregnancy, the weight gain during the pregnancy, smoking, experiences of previous abortions and previous miscarriage, distance to the closest hospital), delivery characteristics (newborn head circumference, monitoring of fetal heartbeat, weekend delivery, night shift delivery, antibiotics during labor), hospital characteristics (the ratio of femoral neck fractures treated within two days), local health authority, month, and year fixed effects. The standard errors are clustered at the hospital, shift, and weekdays level. *Mean of Y* and *SD of Y* report the dependent variable's sample mean and standard deviation, respectively. *Observations* report the number of observations included in the estimation. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9
Effect of a full team on additional outcomes. RDD Reduced form results on the uncomplicated vaginal delivery sample.

	Breastfeeding (1)	No Kristeller (2)	Spontaneous afterbirth (3)	No oxytocin (4)	No prostaglandins (5)	No amniorrhexis (6)
Panel A - Discontinuity sample						
Above threshold	−0.0814***	0.0305*	−0.0120	−0.0281***	−0.0208	−0.0159*
	(0.014)	(0.016)	(0.048)	(0.010)	(0.016)	(0.009)
Mean of Y	0.134	0.923	0.115	0.901	0.854	0.961
SD of Y	0.341	0.266	0.319	0.299	0.354	0.194
Observations	23,435	23,435	23,435	23,435	23,435	23,435
Panel B - Sample 250–749 births						
Above threshold	−0.1025***	0.1207***	−0.0830**	−0.0381***	−0.0067	−0.0210*
	(0.020)	(0.024)	(0.041)	(0.014)	(0.024)	(0.011)
Mean of Y	0.135	0.917	0.101	0.913	0.839	0.973
SD of Y	0.342	0.277	0.302	0.282	0.368	0.162
Observations	9194	9194	9194	9194	9194	9194
Panel C - Sample 750–1249 births						
Above threshold	−0.0133	−0.0672***	0.2108***	−0.0622**	−0.0699***	−0.0601***
	(0.016)	(0.013)	(0.066)	(0.025)	(0.021)	(0.017)
Mean of Y	0.133	0.927	0.123	0.893	0.862	0.953
SD of Y	0.340	0.260	0.329	0.309	0.345	0.211
Observations	14,241	14,241	14,241	14,241	14,241	14,241

Notes: Each column presents the reduced form estimates for Eq. (3). Panel A shows results from the sample, including all deliveries in maternity units that record yearly births within the ±250 bandwidth around the thresholds of 500 and 1000. In Panel B, we include all deliveries in maternity units that record yearly births within the ±250 bandwidth around the 500 threshold (250–749). In Panel C, we include all deliveries in maternity units that record yearly births within the ±250 bandwidth around the 1000 threshold (750–1250). In all specifications, we include the following set of controls: the difference (and its square) between the total number of births at the hospital-year level and the closest threshold (500 or 1000); mothers' characteristics (age, nationality, education level, employment status, marital status of the mother, whether the woman is at her first delivery, delivery week, whether the woman experienced hospitalizations during the pregnancy, the weight gain during the pregnancy, smoking, experiences of previous abortions and previous miscarriage, distance to the closest hospital), delivery characteristics (newborn head circumference, monitoring of fetal heartbeat, weekend delivery, night shift delivery, antibiotics during labor), hospital characteristics (the ratio of femoral neck fractures treated within two days), local health authority, month, and year fixed effects. The standard errors are clustered at the hospital, shift, and weekdays level. *Mean of Y* and *SD of Y* report the dependent variable's sample mean and standard deviation, respectively. *Observations* report the number of observations included in the estimation. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

CRediT authorship contribution statement

Marina Di Giacomo: Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Massimiliano Piacenza:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Luca Salmasi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Gilberto Turati:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

We declare no competing interests.

Acknowledgements

We wish to thank Gabriella Conti, Janet Currie, Carlo Devillanova, Debora Di Gioacchino, Simone Ghislandi, Marit Hinnosaar, Claudio Lucifora, Joan Madia, Francesco Moscone, Cheti Nicoletti, Pau Olivella, Claudio Piga, David Slusky, Odd Rune Straume, Rasmus Wiese, and all seminar participants at the *2023 Economics of the Healthcare Workforce Workshop* (University of Surrey, UK), the *12th Meeting of the Network on Economics of Regulation and Institutions* (University of Genova), the *2022 Meeting of the European Public Choice Society* (University of Minho, Braga), the *26th Conference of the Italian Health Economics Association* (Bocconi University, Milan), the *62nd Conference of the Italian Economic Association* (Università Politecnica delle Marche), the *33rd Conference of the Italian Society of Public Economics* (University of Bari), the *4th Dondena Workshop on Public Policy* (Bocconi University, Milan), the *34th National Conference of Labour Economics* (University of Piemonte Orientale, Novara), the *4th Edition of the American-European Health Economics Study Group* (WU Wien), the *10th Meeting of the Network on Economics of Regulation and Institutions* (Università Cattolica del Sacro Cuore, Rome), the *30th Conference of the Italian Society of Public Economics* (University of Padova), for their helpful remarks on preliminary drafts of this paper. We also thank Maria Maspoli, Paola Ghiotti, and Gloria Prina at the Assessorato Regionale alla Sanità, Regione Piemonte, for giving us access to the CEDAP microdata. The comments of the Executive Guest Editor, Giuseppe Moscelli, and of three anonymous referees are gratefully acknowledged. All remaining errors are our own.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.labeco.2025.102760.

Data availability

The authors do not have permission to share data.

References

ACOG - American College of Obstetricians and Gynecologists, 2017. Guidelines For Perinatal Care, 8th Edition.
Aiken, L.H., Cimiotti, J.P., Sloane, D.M., Smith, H.L., Flynn, L., Neff, D.F., 2011. The effects of nurse staffing and nurse education on patient deaths in hospitals with different nurse work environments. Med. Care 49 (12), 1047.
Almond, D., 2006. Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the Post-1940 US. population. J. Polit. Econ. 114 (4), 672–712.
Almond, D., Chay, K.Y., Lee, D.S., 2005. The costs of low birth weight. Q. J. Econ. 120 (3), 1031–1083.
Almond, D., Doyle, J.J., Kowalski, A., Williams, H., 2010. Estimating marginal returns to medical care: evidence from at-risk newborns. Q. J. Econ. 125 (2), 591–634.
Amaral-Garcia, S., Nardotto, M., Propper, C., Valletti, T., 2022. Mums go online: is the internet changing the demand for health care? Rev. Econ. Stat. 104 (6), 1157–1173.
Angrist, J.D., Lavy, V., 1999. Using Maimonides' Rule to estimate the effect of class size on scholastic achievement. Q. J. Econ. 114 (2), 533–575.
Avdic, D., 2016. Improving efficiency or impairing access? Health care consolidation and quality of care: evidence from emergency hospital closures in Sweden. J. Health Econ. 48, 44–60.

Avdic, D., Lundborg, P., Vikström, J., 2019. Estimating returns to hospital volume: evidence from advanced cancer surgery. J. Health Econ. 63, 81–99.
Avdic, D., Lundborg, P., VikstrÖm, J., 2024. Does health-care consolidation harm patients? Evidence from maternity ward closures. Am. Econ. J.: Econ. Policy 16 (1), 160–189.
Baicker, K., Chandra, A., Skinner, J., 2012. Saving money or just saving lives? Improving the productivity of U.S. health care spending. Annu Rev. Econom. 4, 33–56.
Bartel, A.P., Beaulieu, N.D., Phibbs, C.S., Stone, P.W., 2014. Human capital and productivity in a team environment: evidence from the healthcare sector. Am. Econ. J.: Appl. Econ. 6 (2), 231–259.
Batey, N., Henry, C., Garg, S., Wagner, M., Malhotra, A., Valstar, M., Smith, T., Sharkey, D., European Society for Paediatric Research (ESPR) Neonatal Resuscitation Section Writing Group, 2022. The newborn delivery room of tomorrow: emerging and future technologies. Pediatr. Res. https://doi.org/10.1038/s41390-022-01988-y.
Berta, P., Martini, G., Piacenza, M., Turati, G., 2020. The strange case of less C-sections: hospital ownership, market concentration, and DRG-tariff regulation. Health Econ. 29 (S1), 30–46.
Bharadwaj, P., Løken, K.V., Neilson, C., 2013. Early life health interventions and academic achievement. Am. Econ. Rev. 103 (5), 1862–1891.
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. Q. J. Econ. 122 (1), 409–439.
Bloom, N., Propper, C., Seiler, S., Van Reenen, J., 2015. The impact of competition on management quality: evidence from public hospitals. Rev. Econ. Stud. 82 (2), 457–489.
Card, D., Fenizia, A., Silver, D., 2023. The health impacts of hospital delivery practices. Am. Econ. J.: Econ. Policy 15 (2), 42–81.
Carrillo, B., Feres, J., 2019. Provider supply, utilization, and infant health: evidence from a physician distribution policy. Am. Econ. J.: Econ. Policy 11 (3), 156–196.
Chan, D., 2016. Teamwork and moral hazard. J. Polit. Econ. 124 (3), 734–770.
Chan, D., 2018. The efficiency of slacking off: evidence from the emergency department. Econometrica 86 (3), 997–1030.
Chan D., Y. Chen (2022). The productivity of professions: evidence from the emergency department. NBER Working paper No. w30608.
Chandra, A., Staiger, D.O., 2007. Productivity spillovers in healthcare: evidence from the treatment of heart attacks. J. Polit. Econ. 115 (1), 103–140.
Chandra, A., Staiger, D.O., 2020. Identifying sources of inefficiency in healthcare. Q. J. Econ. 135 (2), 785–843.
Chandra, A., Finkelstein, A., Sacarny, A., Syverson, C., 2016. Productivity dispersion in medicine and manufacturing. Am. Econ. Rev. Pap. Proc. 106 (5), 99–103.
Conti, G., M. Hanson, H. Inskip, S. Crozier, C. Cooper, K. Godfrey (2020). Beyond birthweight: the origins of Human capital. IZA Discussion Paper No. 13296.
Cook, A., Gaynor, M., Stephens, M.Jr, Taylor, L., 2012. The effect of a hospital nurse staffing mandate on patient health outcomes: evidence from California's minimum staffing regulation. J. Health Econ. 31 (2), 340–348.
Currie, J., MacLeod, W.B., 2008. First do No harm? Tort reforms and birth outcomes. Q. J. Econ. 123 (2).
Currie, J., MacLeod, W.B., 2017. Diagnosing expertise: human capital, decision making, and performance among physicians. J. Labor. Econ. 35 (1), 1–43.
Daysal, N.M., Trandafir, M., van Ewijk, R., 2019. Low-risk isn't no-risk: perinatal treatments and the health of low-income newborns. J. Health Econ. 64, 55–67.
Di Giacomo, M., Piacenza, M., Siciliani, L., Turati, G., 2017. Do public hospitals respond to changes in DRG price regulation? The case of birth deliveries in the Italian NHS. Health Econ. 26 (s2), 23–37.
Di Giacomo, M., Piacenza, M., Siciliani, L., Turati, G., 2022. The effect of co-payments on the take-up of prenatal tests. J. Health Econ. 81, 102553.
Doyle, J.J., Ewer, S.M., Wagner, T.H., 2010. Returns to physician human capital: evidence from patients randomized to physician teams. J. Health Econ. 29, 866–882.
Einav, I., A. Finkelstein, N. Mahoney (2022). Producing health: measuring value added of nursing homes. NBER Working Paper Series, No. 30228.
Evans, W., Kim, B., 2006. Patient outcomes when hospitals experience a surge in admissions. J. Health Econ. 25, 365–388.
Facchini, G., 2022. Low staffing in the maternity ward: keep calm and call the surgeon. J. Econ. Behav. Organ. 197, 370–394.
Figlio, D., Guryan, J., Karbownik, K., Roth, J., 2014. The effects of poor neonatal health on children's cognitive development. Am. Econ. Rev. 104 (12), 3921–3955.
Friedrich, B., M.B. Hackmann (2017). The returns to nursing: evidence from a parental leave program. NBER Working Paper No. w23174.
Gaynor, M., Laudicella, M., Propper, C., 2012. Can governments do it better? Merger mania and hospital outcomes in the English NHS. J. Health Econ. 31, 528–543.
Gaynor, M., Seider, H., Vogt, W., 2005. The volume-outcome effect, scale economies, and learning-by-doing. Am. Econ. Rev. Pap. Proc. 95, 243–247.
Gruber, J., Kim, J., Mayzlin, D., 1999. Physician fees and procedure intensity: the case of cesarean delivery. J. Health Econ. 18, 473–490.
Gruber, J., Kleiner, S.A., 2012. Do strikes kill? Evidence from New York State. Am. Econ. J.: Econ. Policy 4 (1), 127–157.
Gruber, J., Wise, D., 2002. An international perspective on policies for an ageing society. In: Altman, S.H., Shactman, D.I. (Eds.), Policies For an Ageing Society. The Johns Hopkins University Press, Baltimore, pp. 34–62.
Grünebaum, A., McCullough, L.B., Brent, R.L., Arabin, B., Levene, M., Chervenak, F., 2015. Justified skepticism about Apgar scoring in out-of-hospital birth settings. J. Perinat. Med. 43 (4), 455–460.
Hansen, K., 2016. Breastfeeding: a smart investment in people and in economies. Lancet 387 (10017), 416.
La Forgia, A., 2022. The impact of management on clinical performance: evidence from physician practice management companies. Manage Sci. 69 (8), 4646–4667.
Lee, T., Propper, C., Stoye, G., 2019. Medical labour supply and the production of healthcare. Fisc. Stud. 40 (4), 621–661.
Leinweber, J., Fontein-Kuipers, Y., Karlsdottir, S.I., Ekström-Bergström, A., Nilsson, C., Stramrood, C., Thomson, G., 2023. Developing a woman-centered, inclusive definition of positive childbirth experiences: a discussion paper. Birth 50 (2), 362–383.
Lin, H., 2014. Revisiting the relationship between nurse staffing and quality of care in nursing homes: an instrumental variables approach. J. Health Econ. 37, 13–24.
Maibom, J., Sievertsen, H., Simonsen, M., Wust, M., 2021. Maternity ward crowding, procedure use, and child health. J. Health Econ. 75, 102399.
Marks, M., Choi, M., 2018. Baby boomlets and Baby health: hospital crowdedness, hospital spending, and infant health. Am. J. Health Econ. 5, 1–48.
Matsudaira, J.D., 2014. Government regulation and the quality of healthcare: evidence from minimum staffing legislation for nursing homes. J. Hum. Resour. 49 (1), 32–72.
McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: a density test. J. Econom. 142 (2), 698–714.
Mesman, R.G., Westert, P., Berden, B.J.M.M., Faber, M.J., 2015. Why do high-volume hospitals achieve better outcomes? A systematic review about intermediate factors in volume–outcome relationships. Health Policy 119 (8).
Ministero della salute, 2010. Gli standard per la valutazione dei punti nascita. Technical report. https://www.salute.gov.it/imgs/C_17_pagineAree_4483_listaFile_itemName_2_file.pdf.
Nicholson, S.C., Propper, 2011. Medical workforce. In: Pauly, M.V., Mcguire, T.G., Barros, P.P. (Eds.), Handbook of Health Economics, Handbook of Health Economics, 2. Elsevier, Amsterdam, pp. 873–925.
O'Donnell, C.P., Kamlin, C.O.F., Davis, P.G., Carlin, J.B., Morley, C.J., 2006. Interobserver variability of the 5-minute Apgar score. J. Pediatr. 149 (4), 486–489.
OECD, 2015. Fiscal Sustainability of Health Systems. OECD Publishing, Paris.
OECD, 2018. Health at a Glance: Europe 2018. State of Health in the EU Cycle. OECD Publishing, Paris.
OECD, 2024. Fiscal Sustainability of Health Systems: How to Finance More Resilient Health Systems When Money Is Tight? OECD Publishing, Paris. https://doi.org/10.1787/880f3195-en.
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. J. Hum. Resour. 43 (1), 88–138.
Perucca, G., Piacenza, M., Turati, G., 2019. Spatial inequality in access to healthcare: evidence from an Italian Alpine region. Reg. Stud. 53 (4), 478–489.
Phibbs, C.S., Mark, D.H., Luft, H.S., Peltzman-Rennie, D.J., Garnick, D.W., Lichtenberg, E., McPhee, S.J., 1993. Choice of hospital for delivery: a comparison of high-risk and low-risk women. Health Serv. Res. 28 (2), 201–222.
Piacenza, M., Turati, G., Vannoni, D., 2010. Restructuring hospital industry to control public health care expenditure: the role of input substitutability. Econ. Model. 27 (4), 881–890.
Rachet-Jacquet, L., Gutacker, N., Siciliani, L., 2021. Scale economies in the health sector: the effect of hospital volume on health gains from hip replacement surgery. J. Econ. Behav. Organ. 190, 704–729.
Raja, C., 2023. How do hospitals respond to input regulation? Evidence from the California nurse staffing mandate. J. Health Econ. 92.
Rogowski, J.A., Staiger, D., Patrick, T., Horbar, J., Kenny, M., Lake, E.T., 2013. Nurse staffing and NICU infection rates. JAMA 167 (5), 444–450.
Shrank, W.H., Rogstad, T.L., Parekh, N., 2019. Waste in the U.S. Health Care system: estimated costs and potential for savings. JAMA 322 (15), 1501–1509.
Skinner, J., Staiger, D., 2015. Technology diffusion and productivity growth in health care. Rev. Econ. Stat. 97 (5), 951–964.
Silver, D., 2020. Haste or waste? Peer pressure and productivity in the emergency department. Rev. Econ. Stud. 88 (3).
Turati, G., 2013. The Italian *Servizio Sanitario Nazionale*: a renewing tale of lost promises. In: Costa-Font, J., Greer, S.L. (Eds.), Federalism and Decentralization in European Health and Social Care. Palgrave Macmillan, London.
Vento, M.P.Y., Cheung, M., Aguar, 2009. The first golden minutes of the extremely-low-gestational-age neonate: a gentle approach. Neonatology 95 (4), 286–298.
World Health Organization, 2018. WHO Recommendations: Intrapartum Care For a Positive Childbirth Experience. World Health Organization.
Wiegerinck, M.M.J., Eskes, M., van der Post, J.A.M., Mol, B.W., Ravelli, A.C.J., 2020. Intrapartum and neonatal mortality in low-risk term women in midwife-led care and obstetrician-led care at the onset of labor: a national matched cohort study. Acta Obstet. Gynecol. Scand. 99, 546–554.
Zweifel, P., Felder, S., Meiers, M., 1999. Ageing of population and health care expenditure: a red herring? Health Econ. 8 (6), 485–496.