PHYSICIANS’ TREATMENT BEHAVIOUR:

DRIVERS, INCENTIVES, AND MOTIVATIONS

Thesis presented for the degree of
DOCTOR OF PHILOSOPHY
in the Faculty of Population Health Sciences

Field of study:
Economics

PAULA DE SOUZA LEAO SPINOLA
[redacted]

Institute for Global Health
UNIVERSITY COLLEGE LONDON
2023

DECLARATION

I, Paula de Souza Leao Spinola confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

DEDICATION

I dedicate this thesis to my family, especially my mother and father, who in similar and complementary ways have offered me the support to begin, endure, and finish this cycle.

To my dear father, who offered me the best education I could have wished for. You provided me with much more than just investments. You made me feel protected – and it was this sense of protection that many times allowed me to take risks, look forward, and focus on what I had to. Dad, you taught me many lessons – and to stay connected to my heart was probably the most important of them. I am deeply proud to have a father that carries your soul, your genuine generosity, and your deep appreciation for human connection.

To my dear mother, you have been my most generous tree in life and an example of strength, commitment, and care. You taught me about responsibility and determination, without which I would not have been able to finish this demanding cycle. I admire your capacity to use your intelligence and innate dedication to everything you do. Your ability to excel in all tasks still takes my breath away. That combined with your loyalty, unconditional love and relentless determination make you who you are – and me, a very fortunate daughter.

To my sister, for having remarkably anticipated the key stages of life to me. This was probably the first major life cycle which I went through without having had your example beforehand. It has been challenging, but I managed!

To tia Kiki, for being much more than a backup mother but my third parent in its own entirety. I was granted the luck to have been born with the reassurance of having a second mother, and I feel even luckier to become closer and closer to you as I age.

To tio Marcio, for being the most academically-mind person in the family and for sharing his knowledge (in all matters) again and again. I admire you not only for your intellect, but also for your ethic and good intentions in life.

To Gabal and Idi, for having witnessed my efforts since a very early age and for providing me with the limitless opportunity to alternate study with laughter and joy. You will always represent to me my extended family, a mix between parents and siblings and, above all, friends that life has gifted me with since birth.

To my grandmother, for her gentleness and remarkable capacity to, effortlessly, be a woman ahead of her time. You are timeless.

ACKNOWLEDGEMENT

First, I would like to thank and acknowledge my primary supervisor and mentor, Marcos Vera-Hernandez, who has not only shared his knowledge with me, but also supported my personal development and gave me encouragement to take my own steps. I am deeply grateful for your time, kindness, and constructive advice. I appreciated the chance to engage in conversations with you about empirical estimates, research strategies, available literature, Brazilian institutional settings, past and future projects, personal challenges, and career prospects. You opened many doors to me and helped me cross all of them. I am certainly finishing this cycle as a very different person who began it and I have absolutely no doubt that your support was essential for my progress.

I am also grateful to Aureo de Paula, my auxiliary supervisor. You have always been available when I needed and offered knowledge at the frontier. Your efficiency, rapid responses, and ease to dive in and out of technical details while staying connected with the overall scope are impressive.

I would also like to thank colleagues from the Institute for Global Health (IGH) and Institute for Fiscal Studies (IFS). At IGH, I learnt from multidisciplinary teams of professionals concerned about health topics and acquired teaching experience which allowed me to take a break from research while developing complementary skills. At IFS, I was given the opportunity to engage in a UK policy-oriented project and leant from expert economists engaged in applying their skills to a broad scope of social issues.

From the IFS, I would also like to especially thank Britta Augsburg. While it has been challenging to work on a parallel project while finishing my PhD, it was certainly less stressful with your help. You have always been understanding and easy to talk to, while being committed to work and deadlines. I admire your flexibility, ability to improvise and think “out of the box”, as well as your ever-present enthusiasm and optimism.

Finally, I would like to thank Rudi Rocha, my former MSc supervisor whose help and support was a determinant in my career. I still remember our talk in your office at UFRJ when you first mentioned about the richness of Brazilian health records and limitless research possibilities. And here I am, working with them until this day. Our encounter was certainly a stepping stone in my journey and brought healthcare to the forefront of my ambitions!

And, of course, I could not end this without thanking and acknowledging the support of my friends. A special thank you to Meag, Nicole, Ana, Rolando, Ines, Fernanda, Alice, Lele, Giulia, Luiza, Isabella, and Helena. Finally, a deep thank you to Jeanne who helps me keep an open heart and mind towards all that life has to offer.

ABSTRACT

There is evidence of large variation in medical treatments delivered across regions as well as across physicians within the same clinical environment which is not explained by patients' characteristics. This thesis uses rich data from Brazil to empirically investigate three questions related to the determinants of physicians' treatment behaviour in the hospital setting.

First, we shed light on the role of peers in shaping physicians' practice styles. We examine whether physicians' use of hospital resources responds to fluctuations in health spending of their nearby colleagues acting in the same medical specialty. We find that physicians incorporate to their own spending roughly *half* of the observed variation in peers' average spending during the preceding 30 days. Peers' gender composition is also found to be a strong determinant of physicians' behaviour. Working around a higher proportion of female doctors causes physicians to take less resource-intensive decisions.

Next, we assess the impacts of a federal policy that rationed compensation in the relative use of C-sections. Although financial (dis)incentives were introduced at the hospital level and didn't directly affect physicians' remuneration, C-section use decreases markedly in municipalities facing high constraints from the policy. Findings that such decreases were followed by health improvements provide evidence that unjustified (and harmful) C-sections were being systematically conducted in municipalities with high C-section rates prior to the policy.

The third question is related to birth timing manipulation around days characterised as being inconvenient for women to deliver, for physicians to work, and/or related to changes in the quality of hospital services. We observe that, while physicians in the private sector accommodate mothers' preferences as well as their own demand for leisure and schedule constraints, manipulation in public hospitals is more limited and occurs, to some extent, in response to the risk profile of births and quality of service delivery.

IMPACT STATEMENT

My doctoral research uses rich data from Brazil to study topics related to the treatment behaviour of doctors in the hospital setting. The evidence generated by this thesis implies that, although there is substantial variation in the way care is delivered by different health providers, the practice styles of physicians are, to some extent, malleable. This study contributes to academic research by presenting empirical findings that contradict the hypothesis that physicians' styles are largely fixed and, subsequently, argues that there is considerable scope for the implementation of policies that nudge physicians to behave in ways that are judged to be more beneficial to patient welfare.

Compelling evidence points to clinical environment as an important determinant of physicians' medical decisions. First, this thesis reveals the significant role of peers in shaping physicians' treatment behaviour. Policymakers could, therefore, consider altering the composition of medical teams as a means to induce physicians to adjust their practice styles. These types of policies are reasonably feasible in contexts where there is a single payer or in systems where the provision of care is not too fragmented. It is also safer to implement such policies in hospital departments where there is larger scope for social learning – these are usually places where decisions are discussed in groups and services are not typically delivered under pressure, such as diagnostic assessments and chronic disease management.

Second, this study shows that physicians' decisions are sensitive to institutional-level incentives/constraints. When incentives/constraints are uniform regardless of patients' characteristics despite treatment returns varying across patient groups, implementing them at the institutional rather than the physician level tends to be a more desirable alternative. This is because physicians are expected to be more responsive to the different medical needs of individual patients if they do not directly face rewards/penalties from their clinical decisions. While centralised governmental actions should be considered whenever physicians tend to systematically make inferior decisions, it would be prudent to simultaneously have in place clinical guidelines recommending the most appropriate treatment by patient type according to available medical knowledge.

Finally, this thesis documents that the role of non-medical motivations in treatment decisions is more salient in the private healthcare system. Regulators should consider monitoring the private sector more closely given that the non-medical factors at play may have negative consequences to patient health and efficiency in care provision. While patients' preferences are expected to be more prevalent in the private sector, patients should be informed in case their preferences produce suboptimal decisions. Besides, policies should be formulated to reverse supply-side incentives that result in decisions that are not in the best interest of patients. Finally, understanding systematic differences in care patterns across the public and private systems is fundamental for the design of policies aimed at decreasing inequalities in healthcare. While care delivered in the public sector tends to be guided by disease severity to a greater extent, there could be situations where routinely available services in private hospitals with high returns to patient outcomes are largely inaccessible in the public system.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Using very rich data from Brazil, *this thesis aims to investigate the malleability of physicians' clinical patterns in response to different external factors*. It examines the influence of peer exposure, institutional incentives, and individual motivations on physicians' clinical decisions, seeking to illuminate the extent to which external elements shape physicians' choices during decision-making processes. Understanding whether physicians adapt their treatment decisions to external changes is crucial, as it provides policymakers with valuable insights into the factors influencing clinical decision-making. These insights have the potential to shape policies that induce physicians to make more optimal decisions, aiming at enhancing patient welfare.

There is incredibly large variation in how doctors treat the same type of patient with a given medical problem. A discussion on the factors driving differences in physicians' clinical behaviour is provided in Chapter 2, along with a review of studies underlying these determinants. Physicians who systematically choose different treatment alternatives tend to differ in terms of underlying preferences, beliefs of treatment appropriateness, or intrinsic skills (in either issuing diagnoses or conducting medical interventions). Most studies relied in physicians' fixed effects to identify practice styles, thus assuming their behaviour is largely fixed. *This thesis challenges this assumption by demonstrating that institutional incentives, peer groups, and treatment timing influence physicians' treatment dynamics*. This is a crucial consideration as it indicates that there is room for policymakers to influence clinical decisions if they deem it beneficial. The final section of Chapter 2 provides more detail on the aims, objectives, and main motivation of this thesis. The subsequent three chapters of the thesis comprise different analyses, with each one dedicated to investigating one of these three factors.

Existing evidence documents that shifting practice environments triggers physicians to adjust their treatment behaviour (Avdic et al., 2023; Doyle & Staiger, 2022; Molitor, 2018). Given that changes in practice environment entail altering many aspects likely to influence clinical decisions (i.e., institutional incentives, infrastructure, colleagues), these findings are usually unable to disentangle the contribution of peer effects. In the next chapter of this thesis, we empirically investigate whether peer interaction is an important channel through which physicians adjust their practice styles. Differences in physicians' clinical practice may be a result from having formed/acquired distinct preferences, beliefs, and skills throughout dissimilar educational and career trajectories (e.g., medical school training, mix of patients seen) instead of innate tastes and hard-to-change tendencies. By working together, physicians have the opportunity to update factors that predict clinical decision-making. This could be mediated by formal/informal collaboration, information exchange, or simply by observing others with dissimilar treatment styles. The latter could occur if physicians detect higher returns from other clinical pathways chosen by their peers or if they feel pressured to comply to social norms. Behavioural changes spurred by social pressure could still lead to knowledge acquisition through the experimentation of new treatment routes.

To answer the question "Does exposure to peers of different styles cause physicians to update their own treatment behaviour patterns?", Chapter 3 uses data on the entire network of approximately 200,000 physicians providing hospital services within the Brazilian public health system between July/2012 and December/2019. Episode-level data on physician activity is linked to Medical Council registries containing their individual characteristics. We measure treatment styles with information on the total costs of conducted hospitalizations and define peers as physicians of similar medical expertise who work in the same hospital-month. Because physicians may work simultaneously in multiple hospitals and be registered in several medical specialties, physicians' networks do not perfectly overlap in both dimensions (i.e., place of work & medical specialty). We exploit exogenous variation generated from this feature by instrumenting, in a linear-in-means model, peer cost with fixed characteristics of peers of peers who have never worked in the same hospital nor shared a medical specialty with the focal physician. In the presence of peer effects, a shock to the composition of peers of peers would trigger peer behaviour to change which, in turn, would affect the behaviour of the focal physician. Our results point to two important findings. First, we find economically significant behavioural spillovers in physicians' use of hospital inputs. Doctors incorporate to their own spending roughly *half* of the observed variation in the average cost of hospitalizations conducted by their peers during the preceding 30 days. Second, our findings point to peer gender composition as a strong determinant of physicians' behaviour. Changes in the proportion of female peers has a direct effect (i.e., conditional on their behaviour) on focal physician's outcome that is generally equivalent, in order of magnitude, to the indirect effect triggered by behaviour spillovers (i.e., focal physician responses to the observed changes in peer behaviour associated with the gender compositional change). More specifically, we show that a marginal increase in the share of female peers (at its sample average) causes physicians to decrease their hospital spending by 8%. This finding is of special relevance given recent trends of increased female representation in traditionally male-dominated medical fields.

In the following two chapter, we turn our attention to hospital care during childbirth. While differences in medical treatments across providers is not necessarily inefficient (as has been extensively argued in Chapter 2), this is unlikely to be the case when we look at choice of method of childbirth delivery. The widespread use of C-section and increasing deviation from recommended rates are indicative that a significant proportion of these procedures are motivated for reasons unrelated to medical need. The literature points to abundant variation in providers' systematic childbirth procedure choice (Card et al., 2023; Currie & Macleod, 2017; Epstein & Nicholson, 2009). Studying ways to disincentivize medically-unwarranted C-sections is particularly important in settings where non-medical factors play a relevant role in childbirth procedure decisions. Latin America and the Caribbean stand out as the region with the highest C-section rate in the world (43%), while Brazil acts a major contributor with 56% of births having been delivered by the surgical procedure between 2010 and 2018 (Betran et al., 2021).

While variations in medical treatments among providers may not necessarily be inefficient, as extensively discussed in the preceding chapters, this is less likely to hold true when considering the choice of childbirth delivery method. There is a clear understanding that unwarranted C-sections are increasingly used. The global proportion of births delivered by C-section has risen significantly from 7.6% to 21% between 1994 and 2021, exceeding the 15% level recommended by the World Health Organization (WHO). In numerous nations, the prevalence of this surgical alternative has surpassed that of vaginal deliveries (Betran et al., 2021; Betrán et al., 2016). Factors such as advancing maternal age, higher shares of women with prior C-sections, and improved procedural safety may account for part of this upward trend (Lancet, 2000). Yet, the rapid increase and deviation from recommended rates strongly suggests that a significant portion of these procedures are driven by factors unrelated to medical need.

Chapter 4 evaluates a national reform in the late 1990s that introduced financial incentives (at the institutional level) for choice of vaginal delivery, the alternative procedure choice. The main feature of the reform was the introduction of a fixed cap to the monthly rate of C-sections that hospitals in the public sector could claim compensation for. We ask: "Did physicians respond to hospital-level financial incentives by reducing C-section use? What were the impacts of the policy for infant health?" Using a differences-in-differences empirical design, we find that municipalities more constrained by the fixed threshold (i.e., those with higher propensity to perform C-sections prior to the policy announcement) experienced significantly larger decreases in C-section likelihood. Decreases were higher among births less likely to be associated with medically justified C-sections (first-order births, younger mothers, single pregnancies) and were accompanied by health improvements at the time of birth as well as during the 365 days following birth. In terms of outcomes at birth, we find that children born in municipalities less likely to perform C-section due to the policy experienced reduced likelihood of low birthweight - which also suggests that at least part of the C-sections eliminated by the policy would have been performed prematurely. In terms of later outcomes, we show that more policy-constrained municipalities (where C-section likelihood dropped more significantly) experienced higher drops in the total number of (public sector) hospitalizations of infants born after (vs before) the policy announcement. These drops in hospitalizations were particularly driven by respiratory disorders, including chronic pulmonary disorders and pneumonia/influenza. The fact that the policy significantly decreased C-section use, as intended, while improving children health is highly informative: not only unnecessary C-sections were prevalent prior to the reform, but also those which were detrimental to patient health. The chapter also presents further analyses to test against alternative hypotheses that would threaten the causal interpretation of our results, such as self-selection of expectant mothers to municipalities less constrained by the threshold and increases in the likelihood of admitting children to the private sector, as a response to the policy.

Finally, Chapter 5 assesses whether non-medical reasons influence the time at which births are delivered by physicians. While the onset of spontaneous labour is expected to be uniformly distributed in time around the final weeks of pregnancy, the exact time of births can be manipulated by the use of medical technology such as C-sections and, to a lesser extent, labour induction. In this chapter, we ask: "Is there manipulation in the timing of births away from days characterised as being inconvenient for mothers and/or physicians? Is manipulation accompanied by changes in the mode of delivery? How does it interact with the risk profile of births?" We study these questions by examining over 20 million births delivered in Brazil during the years between 2006 and 2019 to assess changes in timing around days on which the Brazilian Congress of Obstetricians and Gynaecologists is held (inconvenient to physicians), of inauspicious dates (inconvenient to parents), and of bank holidays (typically inconvenient to both sides, but also times when hospital resources might be scarcer, and risk is higher). We investigate these patterns separately for black and white mothers, and for public and private hospitals. We present evidence of birth timing manipulation in both the private and public sectors, which is substantially more salient among white women delivering in private hospitals. While convenience seems to explain most (if not all) manipulation in the private sector, we argue that at least part of the manipulation in public hospitals reflects medical appropriateness and consequently contributes to reductions in racial disparities in quality of received care. At times of expectedly lower quality of service delivery, manipulation in the public sector is especially targeted at births from black mothers and riskier pregnancies. The same pattern is observed within the same hospital, as their funding becomes more (or less) attached to SUS over the period of our time sample.

The first section of each analytical chapter (Chapter 3, 4, and 5) provides further motivation for the respective question under investigation, briefly describes the adopted methodology and main results, and summarises our contribution to the broad literature. Policy discussions and recommendations of future research can be found in the last section of each one of these chapters. The final chapter of the thesis concludes by wrapping up our main findings regarding the examined supply-side factors guiding medical decisions and suggesting future research avenues.

2 PHYSICIAN TREATMENT STYLES

2.1 Understanding what is behind

It is not unusual for physicians to systematically differ in their treatment decisions for the same type of patient, holding constant other incentives/constraints usually defined at a more aggregate level (e.g., financial incentives, legal protection against litigation, infrastructure constraints). There exists mounting evidence pointing to wide variation in physicians' treatment patterns, which the literature usually refers to as *physicians' practice styles*. In this section, we argue that variation in physician clinical decisions is not surprising. Indeed, this is largely expected due to several factors such as awareness of available evidence on treatment returns, differences in physicians' skills and preferences as well as uncertainty in medical knowledge. While such factors tend to operate simultaneously, we present them separately below for ease of interpretation. Later in the chapter, we summarize the literature.

Physicians may have varying understandings of the patients' underlying health issue. The same patient may be offered different treatments simply because they receive conflicting diagnosis assessments. Patients with the same medical problem who receive the same (accurate) diagnosis may still be treated differently by different doctors as there usually are several treatment alternatives for the same health problem. The best treatment option is likely to vary across patient type. Physicians may differ in their awareness of the available scientific evidence on how treatment success rate interact with patient characteristics (e.g., age, comorbidities, genetic inheritance). In addition to issuing the correct diagnosis and knowing the evidence on heterogeneous treatment effects, physicians also need to identify all these relevant attributes when seeing a patient. Moreover, the same individual patient may face different optimal treatment choices when seen by doctors with different sets of procedural skills.[1] As in any field involving individual behaviour, physicians may be better at performing different procedures and therefore experience distinct returns from the same treatment option. This may be related to their own specialization/experience or be intrinsic to innate abilities. In other words, the returns of a given treatment option may vary not only according to the characteristics of the patient who receives it, but also according to the skills of the provider performing it.

---

[1] This thesis concentrates in hospital care, where treatments involve medical procedures. Procedural skills could also be extended to primary care in case physicians vary in their abilities to provide follow-up care after a treatment recommendation.

Now consider a hypothetical world where all physicians have identical skillsets (in diagnosing and performing medical procedures) as well as same level of awareness of available evidence-based research on treatment options and their heterogeneous returns to different patients. There could still be differences in clinical decisions given that providers may differ in a more subtle margin: their belief on treatment appropriateness. Indeed, inherent uncertainty in medical evidence makes room for differences in opinion to emerge. In areas of care where available scientific evidence doesn't dictate a superior clinical pathway, divergence of opinions is expected and not suboptimal. In contrary, they may be desirable as diversifying among undominated treatment options tend to, on the one hand, be safer for patient health and, on the other hand, expand knowledge on treatment effectiveness. Furthermore, in case of increasing returns from specialization, having physicians specialize in different (undominated) treatment options would be, in theory, pareto efficient.

Yet, physicians may have conflicting beliefs in situations where scientific evidence is highly suggestive in one direction. This could happen if physicians' implicit reference of treatment success rates is inferred from their own accumulated experience. First-hand experience is supposed to bring valuable information to physicians in terms of their own skills in performing the given procedure as well as heterogeneous treatment effects based on patient attributes for which no external evidence is available. Although the additional information would be expected to improve decision-making in a rational world, evidence suggests that physicians tend to overrate the value of their own personal experience.[2] Physician convictions that are strongly determined by their own cumulative success rates could be largely biased if they are based on small sample sizes and/or not properly adjusted by relevant patient attributes.[3] Besides, physician may not measure uncertainty appropriately when forming beliefs and making treatment decisions.[4] There is substantial evidence that individuals are not good at making decisions under uncertainty even when they are

[2] Literature in psychology has shown that patient outcomes tend to be inferior when physicians rely on their clinical judgment instead of strictly complying to guidelines (Grove et al., 2000).

[3] The composition of the patients seen (if not representative of the population suffering from the respective condition) is likely to bias physicians' beliefs about average treatment returns (for a random patient). Small sample biases could arise from physicians inferring returns for a patient group they have only occasionally treated or from junior physicians valuing their own discretion too soon in their careers.

[4] While uncertainty across possible health states should ideally be measured in ranges (e.g., 20-40% probability of death), individual physicians are likely to assume *precise* probabilities (e.g., 30% probability of death). If a physician assumes a precise probability of success for each one of the treatment options, he/she is likely to find a single superior alternative even when there is none. If such point probabilities differ (even slightly) across physicians, we could easily be in a world where different physicians have strong diverging convictions on how to treat a patient when, in reality, there could be several reasonable alternatives.

considered to be experts in the respective field (E. J. Johnson, 1988) – and no reasons why physicians would be an exception.

Let's go one step beyond and consider that physicians have the same medical knowledge and skillset as well as a common understanding that there is no superior treatment choice when caring for a given patient (i.e., no diverging beliefs). Once again, they could still choose different routes of treatment because of differences in individual preferences.[5] Some physicians may choose procedures which are more convenient (e.g., procedures which are easier to schedule, have shorter duration, requires less coordination with other providers/hospital staff), others may be tempted to prescribe treatments that are financially more attractive (especially if paid by a third party, such as patient insurance or employer). Altruistic physicians are more likely to accommodate patient preferences while others might choose the procedure that provides them with gained learning experience. The latter could also be seen as guided by altruism if motivated by the acquisition of skills likely to benefit future patients as opposed to personal interests such as career progression and future wages.

Related to the last point, there is final dimension that deserves attention. Physicians who are altruistic as well as identical in all margins emphasised before (beliefs, skills, and awareness of external evidence) may still diverge in their treatment choices if they have different objective functions in mind. While some may choose to optimize health outcomes of individual patients, others could be more oriented by public health motivations.[6] Although place of work is expected to largely determine the objective function under decision-making processes, physician discretion makes room for individual preferences/tendencies to play a considerable role.

2.2 Literature review

In Appendix A, we review the economics literature showing evidence that physicians have idiosyncratic styles to deliver patient care and that they matter. Overall, accumulated evidence suggests that approximately 50% of variation in medical practice is physician-

---

[5] Physicians' preferences are usually intertwined with beliefs of treatment appropriateness (e.g., conservative physicians having beliefs that less invasive alternatives are superior). This paragraph considers preferences conditional on beliefs, for ease of exposition.

[6] If population health is to be considered in individual treatment decisions, two dimensions are expected to be incorporated in the decision-making process. One is cost-effectiveness ("*could the same amount of resources have higher returns to other patients if implemented elsewhere?*"); the other is externalities to society ("*how does the treatment decision at hand affect the health of other individuals?*", e.g. infectious disease containment, future benefits from knowledge spillovers).

specific (Grytten & Sørensen, 2003; Huang & Ullrich, 2023; Phelps, 2000; Phelps et al., 1994; Tu, 2017).[7] Physician practice style is usually measured in (price-adjusted) medical spending (costs of treatment, after accounting for underlying health condition), except for diagnostic-specific analysis where specific treatment options can be assessed (e.g., C-section *vs* normal delivery in case of childbirth). While both measures are usually interpreted in terms of intensity of care provided, the former is also referred to as healthcare utilization. In this section, we review research that empirically investigates the underlying factors causing physicians to differ in their treatment decisions, which were informally presented in the beginning of this chapter.

Cutler at al. (2019) shed light on physician preferences/beliefs as strong determinants of physician treatment behaviour. They use separate vignettes from patient and physician surveys from the US to investigate the contributions of physician preferred treatment choice and patient preferences in explaining variations in hypothetical healthcare intensity (measured by expenditures). Consistent with evidence of physicians' practice style, they find that physician type is the main factor behind regional variations in end-of-life care. Based on evidence of lack of association between intensity of care provided and quality measures, they claim that results are not driven by differences in skills. Epstein and Nicholson (2009) proxy physician quality with experience and attended residency programme and interpret that 30% of the across-physician variation in treatment patterns (measured as risk-adjusted C-section) is due to idiosyncratic physicians' factors uncorrelated to their demographics.[8]

In addition to differences in preferences/beliefs, physician quality is known to vary considerably. Recent papers document substantial variation in the likelihood of mortality experienced by patients who are randomly allocated to physicians in primary care services as well as to specialists in the hospital (Ginja et al., 2022; Stoye, 2022). As clarified above, there are two types of skills that determine physician quality: (i) diagnostic skills (ability to issue the accurate diagnosis and identify patient attributes which are relevant to the selection of the most appropriate treatment choice) and (ii) procedural skills (physicians' intrinsic returns to

---

[7] One exception is Kwok (2019) who finds that physicians' contribution to variation in care provision lies around 15%. This is interpreted as a long-term measure given that it nets out potential temporal switching effects.

[8] This figure refers to the difference between the share of explained variance between a regression that includes physician indicator variables and another regression that replaces these indicator variables with observed physician characteristics (i.e., race, gender, specific residency program attended, experience, region).

treatment choice, also known as surgical/manual/practical skills). We begin by reviewing the first skill type.

Currie et al. (2016) present evidence that diagnostic skills affect choice of procedure. They show that physician-specific fixed effect (usually used by the literature to capture practice style) reflect not only physicians' level of "aggressiveness", but also their different abilities to tailor treatment to the needs of individual patients ("responsiveness").[9] The authors estimate physicians' practice style by separately identifying these two dimensions. First, they use a sample of emergency room admissions in accredited teaching hospitals to predict the likelihood that heart attack patients receive the most care intensive alternative based on their characteristics. The estimates are used to construct an index of patient appropriateness. Second, for each physician, the indicator of invasive treatment option is regressed on the constructed patient appropriateness index in addition to an intercept. The intercept and the slope are interpreted as representing, respectively, provider aggressiveness and provider responsiveness to patient attributes. They find substantial variation in the extent to which cardiologists' treatment choice is sensitive to patient characteristics within the same hospital and year. The authors also present evidence that practice style (especially aggressiveness) is quite persistent over time – and thus, would be largely captured by fixed effects.

In another paper, Currie and Macleod (2017) estimate a model of choice of childbirth procedure as a function of measures of provider diagnostic skills (i.e., "responsiveness") and surgical skills (calculated based on observed outcomes of high- and low-risk mothers). To account for measurement errors and endogenous patient allocation, the authors instrument providers' skills with their market-level equivalents. They find that increasing diagnostic skills by 1 standard deviation causes C-section rates to fall 15.5% for women in the bottom half of the risk distribution and to rise 5.5% among women in the high-risk half of the distribution, whereas increases in surgical skills raise the incidence of C-section across the entire distribution (but especially in the bottom). They observe a reduction in the probability of negative health outcome following skill improvement in both dimensions. Grytten et al. (2012) document that the introduction of ultrasound and cardiotocography in maternity care

[9] Providers' level of aggressiveness could be driven by preferences/beliefs that favour the more invasive treatment option, or comparative advantage in performing such treatment option.

in Norway decreased the variation in C-section rates across hospitals by reducing clinical uncertainty about patients' diagnosis of risk factors.

The previous papers present evidence that physicians differ in their ability to respond to relevant patient characteristics when choosing the most appropriate procedure for a certain health condition. The underlying condition that needs treatment may be easy to diagnose (e.g., childbirth), but this is not always the case (e.g., mental health). Besides, for some medical conditions, a prior decision needs making: whether to order a diagnostic test given the observed symptoms. Using data on ordered tests, chosen treatments and subsequent outcomes, Mullainathan and Obermeyer (2022) show that physicians under- and over-test patients when diagnosing heart attack in the emergency departments.[10]

Even when providers do order the relevant physicial tests, they may differ in how they interpret test results and, therefore, issue diagnoses. Chan et al. (2022) document widely different diagnosis rates among US radiologists seeing (randomly allocated) patients at risk of pneumonia. Based on a model that allows for variation in providers' preferences and diagnostic skills, they interpret that the latter explains 39% of the variation in diagnostic decisions. Their estimated model suggests that less skilled radiologists optimally choose lower diagnostic thresholds because they view missing a diagnosis as more costly than misdiagnosis (i.e., incorrectly diagnosing a healthy patient). The same argument is raised by Marquardt (2023), who studies diagnosis of Attention Deficit Hyperactivity Disorder (ADHD).[11] The author estimates that differences in diagnostic (implicit) thresholds across physicians explain 2/3 of the observed variation in ADHD diagnosis. In another paper, she shows that heterogeneity in ADHD diagnosis partly comes from differences in compliance to existing guidelines (Marquardt, 2022).[12] A less recent but commonly cited paper is Song et al. (2010),

---

[10] The authors' machine learning algorithm predict that allocative inefficiency in testing decisions is due to bounded rationality (i.e., physicians' implicit model of risk is too simple) and use of suboptimal weights across relevant symptoms/signs.

[11] Although the decision-making problem is simpler for mental health diagnosis in the sense that it does not involve additional monetary costs (i.e., no physical tests can be ordered), it is highly complex/subjective exactly for the same reason: they cannot be informed by blood tests or medical imaging.

[12] Although non-compliance is problematic if driven by unawareness of existing guidelines or opposing beliefs to latest evidence, it could go all the way to being beneficial because of physicians' superior information about the individual patient under treatment. Physicians, however, are likely to make inferior decisions if they give suboptimal weights to salient information (e.g., overrate the value of their superior information) or do not appropriately account for uncertainty (e.g., make inference from very small samples). In his new book and related academic papers, Manski discusses these issues extensively and proposes recommendations for improvement in physicians' decision-making process based on available information as well as for guideline development (Manski, 2017, 2018, 2019). There is growing causal evidence that patients seen by physicians who diverge from guidelines fare worse (Abaluck et al., 2021; Cuddy & Currie, 2023). Although findings are

who find that observably similar patients who moved to higher-intensity regions in the US received more diagnoses over time than those who moved to lower-intensity regions.

Over-testing and over-diagnosing (after testing) have been shown to be especially prevalent among less skilled providers (Chan et al., 2022; Doyle et al., 2010) – a very intuitive finding: physicians may try to compensate their lower skills by consuming more hospital resources. Consistent with this hypothesis, Silver (2021) and Gowrisankaran et al. (2022) report negative association between physician quality and incurred costs in the emergency department in Canada and the US.

Other studies tried to assess whether differences in physicians' treatment decisions are explained by differences in procedural skills (i.e., physicians' individual ability to perform a given procedure). Chandra and Staiger (2007) provide a theoretical framework consistent with the argument that differences in practice styles *across regions* could, indeed, be explained by specialization in treatment options with highest returns. In their model, treatment choice depends on patient clinical appropriateness as well as heterogeneous *local* returns to treatment. They show that productivity spillover would result in multiple equilibria with different regions specializing in different treatment options. A later paper by the same authors shows that differences in treatment patterns across hospitals could also be driven by hospitals' inaccurate beliefs about their own comparative advantage (Chandra & Staiger, 2020).[13] This provides a framework for the hypothesis that physicians overate the value of their individual contribution to patient care, which is aligned with the argument that uncertainty during medical decision-making is not factored in appropriately.

In their models, Chandra and Staiger (2007, 2020) consider physicians' comparative advantage as given. Comparative advantage could arise from knowledge spillovers (e.g., improved skills spurred by social learning and peer cooperation), or simply be a consequence from selective migration of physicians with fixed skills (e.g., acquired through education or innate ability). Gong (2018) studies how physicians' treatment choices evolve as their skills improve over time. She proposes a dynamic model where physicians not only face higher

informative, effects are likely to be largely dependent on the context (e.g., medical condition, guideline quality) as well as drivers behind non-compliance (e.g., unawareness, valuable superior information, suboptimal problem-solving). Research in this area is highly promising.

[13] The model proposed in the first paper is not able to separate comparative advantage from allocative inefficiency (i.e., treatment decisions influenced by factors other than individual returns to treatments). The two channels are disentangled in the second paper, which uses variation across hospitals instead of across regions.

procedural returns from experience (i.e., "learning-by-doing") but also use their own experience to update their beliefs on treatment effectiveness for different types of patients. Gong shows that while physicians' beliefs are expected to converge over time and, hence, contribute to a reduction in treatment choice variation, learning-by-doing does the opposite: it creates path dependence and reinforces variation.[14] The author, however, does not account for social learning (i.e., learning from peers).

Molitor (2018) investigates the overall influence of practice environments in determining physicians' practice styles. Using variation from cardiologists' migration across regions in the US, he is able to separate physician- and place-specific contributions to treatment choice.[15] Findings suggest that environment accounts for twice as much as physician-specific (time-invariant) factors in explaining practice style. Based on subsample and heterogeneous analyses, the author claims that results are not driven by geographical differences in available technology. He also argues that social learning is unlikely to be the underlying mechanism given that migrants react very fast, a pattern that is more consistent with peer pressure. That said, many other environment-specific factors, such as local system processes and institutional incentives, could be at play. Similar results of rapid adaptation to new environment is shown by Avdic et al. (2023) and Doyle and Staiger (2022) when studying, respectively, physician switches across hospitals and across group practices within the same hospital.[16] After adding hospital fixed effect as an attempt to isolate social factors from hospitals' time invariant factors, the first paper interprets that social factors account for roughly half of the variation. Both papers show evidence of no consequences to patient health, which suggests that treatment intensity induced by the switch may not be productive.

---

[14] Evidence of learning-by-doing in physician decision-making is also shown by Epstein et al. (2016), Facchini (2022), and Lundborg et al. (2021). The literature suggests that, despite long learning curves, initial skill (i.e., performance in a physician's first year of practice) explains most of the variation in physician performance over time. Physicians are likely to become more skilled as they evolve in their career, but skill improvement translates less strongly to improvement in patient health (Epstein et al., 2016; Lundborg et al., 2021). There is also evidence that, like human capital in other occupations, physician procedural skills are likely to depreciate after periods when they are temporarily away from the hospital (Hockenberry & Helmchen, 2014).

[15] More specifically, the author adopts a differences-in-differences approach to study how migrant cardiologists respond to being exposed to different levels of treatment intensity between the new and old practice environments. The key underlying assumption is that physician behaviour is not correlated with unobserved factors which are, in turn, correlated with geographical differences in treatment patterns (e.g., across-region variation in patient unobserved health).

[16] While adaptation to new physical structure is unlikely to be a mediator in Doyle and Staiger (2022)'s findings given that they rely on variation that arises within the same hospital, their results are at higher risk of being driven by physician self-selection to new practice environment (i.e., within-hospital switch is less costly). The authors argue that, if physicians are required to switch affiliation groups before changing their practice style, this would still be variation caused by group affiliation.

2.3 Thesis' contribution

Most studies that identify physicians' practice styles have primarily relied on time-invariant variation, which, by design, fails to capture potential changes in physicians' behaviour over time. The primary goal of this thesis is to investigate the malleability of physicians' clinical patterns under different circumstances, questioning whether these patterns are largely fixed. Specifically, the thesis aims to assess the influence of peers, institutional incentives, and the timing of treatments on physicians' clinical choices within the hospital setting. By examining whether physicians adapt their clinical decisions in response to changes in these factors, we seek to illuminate the extent to which physicians' choices are influenced by external elements during their decision-making process.

In the following three chapters, the thesis explores the role of peers, institutional financial incentives, and motivations associated with treatment timing. Chapter 2 examines whether physicians' hospital spending adjusts in response to fluctuations in observed spending by their peers as well as peers' characteristics. In Chapter 3, we investigate whether decisions regarding childbirth procedures are impacted by changes in hospitals' financial incentives aimed at influencing procedure selection. Lastly, Chapter 4 assesses how specific time periods affect decisions concerning the exact date and method of childbirth procedures. The specific research questions behind these investigations are outlined in detail in the Introduction chapter.

Understanding whether physicians adapt their treatment decisions to exogenous changes around them is critical, as it informs whether physicians are likely to modify their behaviour during their professional careers, when facing different environments and incentives. While many beliefs and preferences may be formed during their educational training, there could still be significant opportunities for physicians to reassess their clinical decisions based on various factors, including where they practice, with whom they work, and when patients require treatment. This thesis explores the influence of institutional features in the workplace, the characteristics and behaviours of peer groups, and the timing of clinical decisions (often linked with inherent physician preferences such as convenience or risk minimization in service delivery) within the decision-making process.

The primary motivation of investigating the factors influencing physicians' decision-making is to provide policymakers with actionable insights on which elements to target when seeking to influence clinical decision-making. This is particularly pertinent given that the reviewed literature indicates that a substantial portion of the observed variation in physicians' treatment patterns is inefficient, thereby highlighting the opportunity for policymakers to design policies that induce physicians to act in a manner that enhances patient welfare.

3 ROLE OF PEERS IN SHAPING PHYSICIANS' TREATMENT STYLES

Practice styles are largely shaped during formal education as well as on-the-job. At the hospital, not only physicians receive continuous training and accumulate practical experience, but they also have the opportunity to exchange and learn from colleagues of varying practice styles. Differences in standards of practice may, indeed, be the result of different educational background and/or accumulated experience instead of hard-to-change intrinsic preferences. There is, therefore, great scope for these interactions to influence providers' practice attitudes.[17] This chapter investigates the role of peers in shaping physicians' practice styles.

The peer effects literature has shown the crucial role of peers in shaping individual behaviour in the most varied settings, even when decisions are known to be harmful such as taking up smoking and engaging in unhealthy lifestyles (Nakajima, 2007; Trogdon et al., 2008). In the context of medical practice, peer effects are of particular interest for society and policymakers. First, the agent making the decision is not the one who directly benefits/suffers its consequences. Although physicians are obviously better placed than patients to choose treatment options, any medical decision consists of a very complex optimization problem. Research shows that physicians frequently fail at accurately accounting for uncertainty when forming beliefs and caring for patients (Manski, 2018, 2019). Being exposed to fellow doctors whose patterns of care differ could serve as an opportunity for physicians to learn how to better incorporate underlying uncertainty in their own decision-making and, consequently, make choices closer to the optimal ones. This is a much more desirable way to speed physicians' capacity building relative to "learning-by-doing" where junior doctors gain expertise sometimes at the cost of human lives. Although social learning and experiential learning are certainly not perfect substitutes, marginally substituting the latter for the former is likely to be beneficial. Nonetheless, peer effects could also have negative consequences if, instead of operating through social learning (i.e., knowledge spillovers), they are driven by pure social pressure. Observing peer behaviour may, for instance, influence physicians to overweight factors that are not determinants of patient outcome, such as personal convenience and institutional incentives. In the end, the extent to

[17] This chapter mostly refers to providers as a synonym for physicians (i.e., those ultimately responsible to perform medical interventions after the patient is admitted to the hospital).

which peer effects are welfare-enhancing depends, on the one hand, whether the care that gets disseminated affects positively patient health and, on the other way, if it is the most cost-effective way to achieve such improvement.

We investigate peer effects in physicians' activity in the hospital setting. We take advantage of rich administrative data from the Brazilian Ministry of Health, rarely used in literature, where we observe the entire network of physicians (and track their activity) in the public healthcare system - a total of approximately 200,000 physicians. Hospital records are then linked to Medical Council registries containing physician-level information on gender, date of birth and educational background (university and residency training). Our analysis focuses on the period between July/2012 and December/2019.

In our empirical framework, physicians are peers if they work in the same hospital-month and share at least one medical specialty. Therefore, in addition to being colleagues, these are physicians who present some similarity in terms of basic medical skills and type of expected treatment decisions (i.e., those for whom information exchange is relevant for their clinical practice), as well as are more likely to engage in social interactions in the workplace given higher proximity (i.e., same hospital department) and common interests/traits (which led them to self-select in the same medical specialty). The main identification challenges are endogenous group formation in both margins (i.e., physician sorting to hospitals and specialties) as well as common shocks to hospitals and specialties. The first challenge refers to the expectation that peers take similar decisions simply because they share similar preferences (which led them to self-select into the same hospitals/specialties and become peers in the first place). The second challenge arises from the fact that, because peers are exposed to the same environment, they are likely to react simultaneously (and similarly) to the same local shocks. These two factors would cause behaviour of nearby colleagues to be positively correlated even in the absence of peer influence. To be able to identify peer effects, therefore, we need to find variation arising from plausible exclusion restrictions (i.e., shocks that solely induce change in the behaviour of peers).

Our setting allows physicians to register in multiple medical specialties and work simultaneously in different hospitals. The exclusion variation we exploit relies on peers who do not overlap in both margins (workplace and specialty). This imperfect overlap allows us to find physicians who are peers with the focal physicians' peers in *other* hospitals and with whom the focal physician does *not* share any medical specialty (hereafter, peers of peers).[18] In an instrumental variable framework, we leverage variation triggered by this source of exclusion restriction to study behavioural spillovers of physicians' medical practice.

Average characteristics of peers of peers (in other medical specialties) are used to instrument peer outcome. This type of instrument was originally proposed by Bramoullé et al. (2009). The main idea is that, in the presence of peer effects, a shock in the composition of peers of peers would induce peer behaviour to change which would, in turn, trigger the focal physician to react by changing her own individual behaviour. The underlying assumption for identification is that, conditional on peer composition, there are no factors directly affecting both the behaviour of the focal physician and the composition of her peers of peers serving in other medical specialties (in which she has never served) at other hospitals (where she has never worked).

We measure physicians' treatment styles using the total costs they generate to the healthcare system based on the national fee schedule. Proxying physicians' styles with healthcare spending has been extensively done by the literature (Grytten & Sørensen, 2003; Kwok, 2019; Tsugawa et al., 2017; Van Parys, 2016). Because more intensive/invasive procedures tend to have higher operational cost, they also tend to be more expensive. This is especially true when there is a single payer determining fixed procedure fees for all providers operating within the system, which is our case. In addition to the cost of the main medical intervention, hospital care involves other costly services which physicians may use to different extents. Some physicians may order more tests, provide additional auxiliary services, and keep patients for longer in the hospital, all of which would contribute to increasing the hospitalization's total cost. Because our measure of cost relies on fixed tariffs for all services provided, differences in total spending across physicians necessarily imply (at least some) variation in service delivery. Besides, in our context, physician decisions are not influenced by personal monetary motivations given that they are typically paid fixed salaries.

Our model consists of a standard linear-in-means model where we regress physician outcome on average peer outcome at the hospitalization level. Peer outcome, our (to-be-instrumented) regressor of interest, is constructed as the average outcome among all

[18] Consider a simple example of two physicians working together: Physician A, who serves as a clinician, and Physician B, who serves as both a clinician and a cardiologist. Physician A's peers of peers would be the group of all cardiologists with whom Physician B works in other hospitals where Physician A has never worked. Physician B, on the other hand, would have an empty set of peers of peers given that her peer's set of medical specialties is a subset of her own.

hospitalizations concluded in the same hospital during the 30 day-period prior to the admission of the focal hospitalization by other physicians in the same specialty (i.e., peers). We control for a large set of fixed effects (year/month, municipality, diagnosis, and specialty) as well as observed characteristics of patient, focal physician, and direct peers. The latter allows us to investigate another type of peer effects: changes on one's behaviour that is triggered by the exposure to certain characteristics of peers, regardless of their respective behaviour. These are known as contextual peer effects. Including peer characteristics into the model also protects us from potential violations of the exclusion restriction because of physician sorting.[19]

We find that physicians' hospitalization costs are affected by observed changes in peers' average cost during the past 30 days. More specifically, focal physicians incorporate to their own spending roughly *half* of the variation in recent peer spending. We also find that physicians' clinical behaviour is affected not only by their own gender but also by the gender of their peers, conditional on peer behaviour. Surprisingly, we find that increasing the proportion of female peers, at the margin, has a direct effect on physicians' behaviour that is generally equivalent, in order of magnitude, to the indirect effect arising from behavioural spillovers (i.e., physicians reacting to the observed change in peer behaviour that is triggered by the initial variation in their gender composition). A marginal increase in the share of female peers (at its sample average) causes physicians to decrease their hospital spending by 8%.

Instrumental variables leveraged from network intransitivity have been previously used across different research areas. In the context of college major choice, De Giorgi et al. (2010) exploit variation from non-overlapping groups of students enrolled across university courses prior to choice of major degree. Looking at firms' decision, Patnam (2015) proposes an instrument based on the fact that firms share the same board director. More recently, two papers encountered intransitivity in social structures by overlaying different networks. To study peer effects in consumption among co-workers, De Giorgi et al. (2019) instrumented co-workers' consumption with characteristics of the co-workers of their spouses. Nicoletti et al. (2018) found intransitive triads after overlapping family and neighbourhood networks

[19] Because peers' characteristics are likely to be correlated with both the characteristics of peers of peers (instrument) as well as those of the focal physician (and, therefore, focal physician's outcome), it is crucial that we control for them.

to examine women's working hours after childbirth.[20] In this chapter, we exploit intransitivity that arises simultaneously in *two* dimensions within the *same* social network.

In the context of medical practice, a similar approach has been adopted by Barrenho et al. (2023) to investigate treatment decisions for a specific medical condition, colorectal cancer. They leverage aggregated variation at the physician-year level, stemming from physician movements across hospitals over time. Our study differs in two significant aspects. Firstly, we exploit the most granular variation that is available, at the hospitalization-level, whereas Barrenho and co-authors aggregate information at the hospital-year level due to low volumes in their particular medical setting. Secondly, we are able to track activity of the *entire* network of physicians within the public health system, while the authors can only observe activity from senior doctors in public hospitals.[21]

Although still limited, existing evidence has raised attention to the significant role of peer effects in healthcare decision-making. Barrenho et al. (2023) show that peer take-up has a meaningful impact on the adoption of colorectal cancer keyhole surgery in the UK. Silver (2021) finds that emergency care physicians immediately incorporate behaviour of their nearby peers in resource use and time spent with patients. Also in the US, Chan (2016) finds that peers affect physician behaviour, through joint monitoring, in the direction of reducing moral hazard in care provision. Agha and Zeltzer (2022) document behavioural spillover in drug prescribing following from changes to financial incentives of individual physicians in the US. While it is difficult to disentangle the particular contribution of peers, Avdic et al. (2023), Doyle and Staiger (2022), and Molitor (2018) show that physicians who switch across practice environments adjust their decision-making to the new work place in the US and Sweeden. Our findings contribute to this growing body of literature, particularly in underscoring the significance of social factors in hospital spending in less developed settings where resources are typically more constrained.

Another strand of the literature to which this study relates concerns the role of team composition. It has been shown that shared past experience (Agha et al., 2022; Bartel et al.,

[20] While all these papers combine a linear in-means models within an IV design based on intransitive social interactions, their models vary in terms of how intransitivity is achieved, temporal space between focal and peer outcomes (as well as time lag between network formation and evaluated outcome), and variables used as instruments (e.g., outcome vs characteristics of peers of peers).

[21] The authors argue that solely observing activities of senior physicians (i.e., consultants) is not problematic as they are considered the primary decision-makers in their context. While supported by evidence from Chan (2021) in the US, indicating that teams assign greater weight to the clinical judgment of senior staff, junior doctors may still exert some influence, albeit to a lesser extent, on team decision-making.

2014; Y. Chen, 2021; Stecher, 2023), medical expertise overlap (Branco et al., 2023), seniority (Bartel et al., 2014; Chan, 2021) and recent training (Berez et al., 2018) of peers are relevant determinants of teams' decision-making and performance. We contribute to this literature by showing that co-workers' gender composition is an important determinant of physicians' treatment decisions. This is consistent with available evidence in the economics literature and other social science domains showing that the dynamics of clinical environments are highly influenced by the presence of female doctors (Cardador et al., 2022; Sarsons, 2017; Wallace, 2014).[22] This might be of special relevance given trends of increasing female representation in medical fields that have traditionally been occupied by males.

It is worth giving special attention to Branco et al. (2023), whose evidence also comes from the Brazilian context. Using the same data source as ours, they show that physicians who provide care within the *same* patient's hospital stay are more productive when they present a higher overlap in terms of medical specialties. A few differences between this paper and ours are worth pointing out. The authors look at teams of all physicians who provide any type of care within a hospital stay where a percutaneous coronary intervention was performed. In our study, we only consider physicians performing the hospital stay's *main* procedure(s) (what Branco and co-authors refer to as proceduralists, e.g., providers who deliver the heart intervention) to study how being exposed to peer hospitalizations affect the way physicians treat their *own* patients – these physicians are supposedly the ones responsible for treatment decision-making during patient stays in the hospital. Besides, we classify as peers all co-workers who share at least one medical specialty (i.e., have some overlap in medical expertise). Evidence from Branco et al. (2023) is reassuring in the sense that physicians seem to be affected by colleagues with similar medical expertise.

A strength of our study is its external validity as we include all hospitalizations for which variation from exclusion restriction is found within the Brazilian public healthcare system network. The drawback from using large amount of data is that our results may be interpreted as being too generic. Yet, we believe to contribute to the existing literature which is usually restricted to very particular settings (e.g., adoption of a specific technology, single

[22] Cardador et al. (2022) and Wallace (2014) argue that females in male-dominant departments are likely to receive less cooperation especially in types of support that are directly related to productivity, such as informational as well as instrumental support (i.e., help in task execution). Sarsons (2017) finds that female surgeons are more penalised from negative performance than their male counterparts, and that negative experience with individual female surgeons are used by physicians to update their beliefs about all female surgeons around them.

hospitals) or specific subsets of the network (e.g., senior doctors, specialists within a medical field). Although we are, in theory, not able to interpret whether peer effects in our context are beneficial or not (in terms of patient health and efficient use of resources)[23], we can say that they exist, are very strong and deserve attention. We expand on potential policy implications in the last section of this chapter.

3.1 Role of peers

3.1.1 Potential channels

Fellow physicians working in the same hospital are likely to come from a wide range of different backgrounds, in terms of education as well as accumulated experience resulting from career trajectory (e.g., places of work since graduation) and mix of patients seen. At the workplace, colleagues have the chance to exchange medical knowledge that stems from awareness of scientific research as well as acquired practical experience. Physicians are, thus, at a position to learn about established medical evidence they were unaware of, as well as to reconsider their interpretations about current knowledge. In doing so, physicians have the possibility to update their understanding of treatment returns as well as improve their abilities to identify relevant patient attributes and issue more accurate diagnosis. All of these could have an impact in physicians' beliefs and, therefore, treatment selection.

In addition to direct information exchange (e.g., discussions about medical research), physicians can gain knowledge by simply observing others treat their respective patients. Physicians are likely to consider experimenting procedures which they observe to be associated with positive patient outcomes. In doing so, they are likely to gain new procedural/manual skills. Besides, witnessing colleagues choose different clinical pathways could lead physicians to become better at incorporating uncertainty in their own clinical judgment.

Finally, working in the same hospital is likely to foster cooperation between fellow physicians. This could happen when physicians (formally) treat the same patient, but also whenever they help each other out (either in information exchange or task-performing). For

[23] Peer effects would be welfare-improving if peers behave according to best practices (i.e., treatment choices aligned with latest scientific evidence, complementary services that help inform most appropriate clinical pathways to follow or those that help reduce risk of complications) and could go all the way to being damaging if physicians incorporate questionable behaviours either in terms of patient health consequences or financial responsibility. The latter includes wasteful spending, when used resources neither helps in health improvement nor risk minimization.

instance, physicians who are unsure about a patient's diagnosis are likely to consult the opinion of others in the department. Hospital departments usually hold team meetings where complex cases are discussed, including diagnosis assessments, complications following a procedure, and the occurrence of medical errors. Many physicians are likely to be involved in the problem-solving at hand given that these hospitalizations tend to last long. In addition to brain-storming discussions, physicians may also cooperate by undertaking specific tasks upon request during the provision of care to patients who they are not formally responsible for. Through effective information exchange and collaboration in patient care, physicians are likely to gain practical experience and update skills that are relevant all along the care pathway.

Summing up, peer interaction provides physicians with the opportunity to update their awareness of medical evidence, beliefs, and skillsets. This is commonly known as social learning. There is another channel through which peer outcome may affect own outcome: social pressure, where the motivation behind convergence in behaviour is wish for peer acceptance and social integration in the workplace. While the former may need time to operate, the latter tends to have more immediate effects. Changes in treatment behaviour initially motivated by social pressure could, however, influence determinants of physician decision-making if they are to be sustained in the long run. As physicians take up new medical procedures, convergence to local standards of care may trigger learning and capacity-building. By experiencing new treatment routes, physicians may also change their beliefs. Ultimately, changes to individual behaviour that sticks in the long run despite no changes to skills/beliefs may persist due to changes in physicians' underlying preferences.[24]

Although peer effects spurred by both channels are likely to interact with the same set of determinants, welfare consequences may not be the same. Learning that is triggered by peer effects are, in theory, expected to be welfare-improving as they bring additional information to physicians' set of accumulated knowledge and capacities. In case of behaviour changes that are initially motivated by social pressure, it is reasonable to expect that physicians are unlikely to engage, at least in the long term, with actions which they realise to

[24] Consider a physician who starts working in a hospital where colleagues turn out to be largely guided by financial incentives. At first, he/she might experience (perceived) pressure to conform and start selecting the usually chosen treatment alternative by others around them. After a while, he/she may realise the personal benefits they extract by behaving in this way and become intrinsically more motivated by personal interests. Preferences could also change for the better (of society) if physicians start working with more resource-responsible colleagues, where they are likely to feel more accountable for their actions.

be harmful to patient health.[25] That said, even in the absence of long-term deterioration to patient health, the second channel could still result in welfare-inferior equilibria if they lead to increases in costs or waiting times despite no increment to patient health.

3.1.2 Literature review

Silver (2021) investigates for (real-time) peer effects in emergency departments. Using variation from physicians shuffling among emergency department teams, he shows that physicians' behaviour, in terms of pace and intensity of care provided, is largely determined by those around them. The author finds rises in patient mortality among providers who reduce time spent with patients and use less hospital resources as a result of working alongside peers who provide hospital care at a faster pace. Based on evidence that physicians who are slower and more resource-intensive are also those who are less productive, he argues that patient health may be compromised if physicians of lower quality are triggered to emulate their higher quality counterparts in highly pressure environments. Chan (2016) shows that social pressure could, on the other hand, be beneficial in environments where joint monitoring causes moral hazard to reduce.

Evidence also exists for peer effects operating through the social learning channel. Exploiting frequent rotation of trainees across teams, Chan (2021) shows that physicians' practice styles change over time as physicians interact with diverse groups of teammates during their training.[26] Results from a structural model estimation imply that there is substantial learning during training, especially when trainees are given a larger stake in team decision-making. These findings are consistent with a large stream of literature pointing to the benefits of teamwork for skill building, efficiency and patient health (Bartel et al. 2014; Chan 2016, among others).

Like this study, a few other papers have leveraged (plausibly) exogenous variation from non-overlapping networks of physicians. A very recent paper is Barrenho et al. (2023), who look at peer effects among senior physicians on innovation take-up of colorectal cancer keyhole surgery in the UK. In their setting, physicians work in a single hospital at any given point in time but may move across hospitals over time. Because they have data since the

[25] Sensible physicians would likely reverse their behaviour towards their past treatment choices (which could, in turn, affect colleagues to take up the more beneficial procedure), or move to hospitals where they don't need to trade-off between patient health and (perceived) social acceptance.

[26] This is based on findings that fixed physician characteristics predict only a small portion of practice variation and that serial correlation grows weaker over time.

innovation introduction, they are able to account for dynamics in physician behaviour. Peer outcome variable, measured as *current* peers' cumulative take-up while working together with the focal physician, is instrumented with characteristics of all *past* peers of peers. Their model also separately identifies effects from working with a larger number of peers and with (endogenously identified) key players. Findings show that, while exposure to key players is more relevant than to larger groups, the effects through the peer effect channel are much more significant. More specifically, one standard deviation (SD) increase in co-worker take-up leads to 3.7 percentual points (p.p.) increase in the focal physician take-up, which is equivalent to 0.13 SD in their sample. They posit that estimates represent the net impact from social learning, peer pressure and norm conformism.

Barrenho et al. (2023)'s largest contribution is the incorporation of effects from exposure to leaders, who have been shown to extensively foment innovation take-up (Agha & Molitor, 2018). However, besides not observing activity for all physicians in the network (which wouldn't be problematic if those physicians don't contribute to team decision-making), their results are based on variation at a relatively aggregate level.[27] While exploiting aggregate-level variation is common in this research area, there are exceptions, such as Yang, Lien, and Chou (2014), who estimate peer effects in the prescription of a new antipsychotic drug in Taiwan using very granular longitudinal data containing repeated interactions between the same patient-physician pair within hospitals over time. Physician outcome is measured as the share of new drug for a given physician-patient pair in a given hospital-month, while peer outcome represents the analogue share averaged among other physicians in the same hospital-month. Estimating first-difference models, they describe that an increase of 10 p.p. in peers' share of the new drug brings a 0.07-0.10 p.p. increase in the physician's own share. Based on findings of stronger effects among more stable peer groups that have existed for longer periods of time, the author suggests that social learning is the underlying mechanism.

In addition to innovation take-up, there is also evidence of peer effects in technology abandonment (Berez et al., 2018). Existing literature also provide evidence of social influence as an important determinant in the use of established technologies (Agha & Zeltzer, 2022; Avdic et al., 2023; Burke et al., 2003; Epstein & Nicholson, 2009) and in medical spending

[27] The authors solely observe the activity of senior physicians in hospital episodes data, yet they are capable of delineating complete employment trajectories. Although some level of aggregation is warranted given their focus on studying uptake as an outcome variable, the decision to aggregate at the yearly rather than monthly level is influenced by the low volume in their setting.

(Doyle & Staiger, 2022). While Agha and Zeltzer (2022) use variation from financial incentives targeted at individual physicians, Avdic et al. (2023) and Doyle and Staiger (2022) exploit variation arising from physician switches across hospitals and across group practices within the same hospital, respectively.

The role of team composition in determining physician decision-making has also been studied. Attention has been directed to the relevance of shared past experience (Agha et al., 2022; Avgerinos & Gokpinar, 2017; Y. Chen, 2021; Stecher, 2023) as well as prior exposure to broader sets of peers (Aksin et al., 2021). In the Brazilian context, Branco et al. (2023) provide evidence that overlap in physician medical expertise is associated with productivity gains (measured in declines in both patient mortality and resource use), most likely due to improved coordination.

3.2 Institutional background

The Brazilian public healthcare system, known as SUS (Sistema Único de Saúde), was inspired by the National Health Services in the United Kingdom and is now one of the largest in the world in terms of coverage (approximately 150 million individuals covered, 75% of the country's population). The services offered by SUS range from simple outpatient care to organ transplantation, all of which are free at the point of use. Even though most of the hospitals operating under SUS are publicly owned, private and philanthropic health facilities might also be contracted by local governments. This chapter will focus on physicians' services provided during hospital stays in SUS.

3.2.1 Payment model and physician attachment

Hospitalizations in SUS are reimbursed based on a national fee schedule. The amount reimbursed comprises of all activity that took place during the hospital stay (including procedures and tests performed, postoperative care while patient is hospitalized, ICU use, and, to some extent, length of hospital stay).

Reimbursement is transferred from the federal government to the local authority that manages health facilities in the area where the hospital is located. The amount of money reimbursed is, therefore, not directly transferred to the hospital where the hospitalization took place. Financial arrangements between the local health authorities and hospitals are defined at the local level (e.g., amount and frequency of payment).[28]

Besides, the amount of money reimbursed must be below a financial ceiling specific to the local health authority. The financial ceiling is negotiated between the federal government and the local health authority every year based on previous year's activity, current fees, as well as plans for the coming year (e.g., planned expansion of a hospital department). Because the claimed amounts tend to be close to (or above) the previously defined financial ceiling, reimbursements end up being largely fixed.

The national fee schedule used to reimburse procedures is issued monthly by the federal government. So far, there has been a widespread understanding that the national tariff is outdated as regards to the fees of many procedures (Machado et al., 2022), e.g. lagging behind inflation. In addition to the fee, the national tariff brings details on procedure compatibility in terms of types of health facility and occupations of health professionals (e.g. medical specialty) eligible to perform it, diagnosis (as of ICD-10) and demographics (e.g. age, gender) of patients who can receive it, among others.

Finally, physicians are usually not paid per services provided. Most doctors are civil servants who have a contract for a fixed number of hours/week (usually 20 or 40 hours/week) and receive a fixed monthly salary regardless of the number and type of procedures they perform. Local governments also have the autonomy to outsource services to private and philanthropic heath facilities (based on the same national tariff in place), and most hospitals are allowed to independently hire physicians – who are usually remunerated by 12- or 24-hour medical shifts. Physicians are allowed to work in multiple hospitals at the same time as well as simultaneously in the public and private sectors.

3.2.2 Patient admission

Patients seen in hospital are either referred from primary care, admitted through the emergency room (if the hospital has one), or transferred from another hospital. Patients can be referred from ambulatory care if diagnostic tests confirm a health problem that requires hospital treatment. Once the referral is put forward, there usually is a consultation with the

---

[28] Local health authorities are ultimately responsible for the financial health of its providers. If, for instance, the federal transfers are not sufficient to cover the hospital bill, they may have to complement the funding with their own local budgets. Although federal reimbursement based on national tariffs may not correspond the exact cost incurred by the hospital, it represents cost from the perspective of the federal government.

relevant team where the procedure is planned and scheduled. Physicians performing the procedure have the discretion to change the original treatment plan. Finally, patients may be admitted through the emergency department in case urgent care is needed.

3.2.3 Physician education and specialty

It is necessary to study 6 years of Medicine undergraduate degree to become a physician in Brazil. Once a degree is completed, physicians need to register in the Regional Medical Council and keep their registration active (by paying an annual fee) to be licensed to offer medical services in the given region. Licensed physicians in Brazil are legally allowed to perform medical treatment in any area of care. Although a formal medical specialty degree is not required by law, hospitals usually hire physicians with further education when seeking specialists.

Physicians may acquire further education in specific medical fields by undertaking formal or informal training. The gold standard in formal specialised education is through medical residency. A physician may undertake multiple residency degrees as some of them have prior degrees as requirements. For instance, a degree in cardiology requires a prior residency degree in either general medicine or general surgery. There are medical residency programmes for the 55 official medical specialties recognised by the Federal Medical Council, of which 24 require the completion of a residency programme in another specialty (Scheffer et al., 2023). Residency programmes last from 2 to 5 years, depending on the medical specialty. There are other post-graduate degrees in specific areas of medical care, which are recognised by the Ministry of Education but are not as prestigious as residency post-graduate programmes. Physicians may also acquire some specialised knowledge through technical short courses and in-hospital training.

3.3 Data

We take advantage of rich granular data that allows us to track physicians' activity across hospitals over time. Our final dataset is the product of linked hospitalization claim records with registries containing individual physicians' demographic characteristics as well as medical education background (for both undergraduate degrees and residency training).

3.3.1 Description of data sources

The databases detailed below are administered by the Brazilian Ministry of Health.

SIH/SUS (*Sistema de Informação Hospitalar do SUS*): administrative data on all publicly funded hospitalizations. The information is publicly available in two separate datasets: RD (*AIH Reduzida*) and SP (*Serviços Profissionais*). Each row of the first dataset corresponds to a hospital episode. The variables include hospital identifier, dates of admission and discharge, diagnosis (as of ICD-10), reason for discharge/closure, total cost (based on the national fee schedule in place at the respective month), patient characteristics (date of birth, gender, postcode of residence), and identification code of the *main* procedure performed. For each of these hospital episodes, the SP dataset contains information on *all* services performed (the main procedure as well as auxiliary services provided) and, for each one of these, the identification code of all health professionals involved and their respective occupations.[29] Since July/2012, the provider's National Health Card (CNS) number was used as the physician identifier.[30] There is no patient identifier. The two datasets can be merged using the hospitalization claim number as key variable. At the time this manuscript was written, the only other (not yet published) paper that has explored the SP dataset is Branco et al. (2023).

CNES (*Cadastro Nacional de Estabelecimentos de Saúde*): contains monthly information on infrastructure and human resources of all accredited hospitals in the country (including those that do not offer any services to SUS).[31] The data comes in different modules, which can be linked together through the health facility CNES's code (which is the same used in SIH/SUS) and the month/year. It comprises general characteristics about the hospital, existing infrastructure, and available services. The module on human resources (CNES/PF) informs, for each registered physician, their name, occupation(s), type(s) of employment contract, as well as numeric individual identifiers: CNS number (same as used in SIH/SP) and their license number issued by the Regional Medical Council, referred to as CRM. Besides, the data contains the individual's encrypted taxpayer identifier, known as CPF (the most reliable identification of Brazilian citizens). The fact that unique CPFs were transformed into distinct encrypted sequences enabled us to generate unique physician identifiers. This is particularly

[29] Auxiliary services include blood test, electrocardiogram, doctor appointments during hospital stay, etc. In case of a childbirth hospitalization, for instance, the delivery method would be the main procedure, and auxiliary services could include newborn doctor appointment, HIV screening, and drug prescription.

[30] Before July/2012, another physician identifier was used. Because this other identifier does not allow us to perform the desired data linkages, we restricted our data sample to begin at this month.

[31] Although we observe all hospitals where a physician is registered to work in a given month (including those operating solely in the private sector), our activity data (SIH/SUS) is restricted to the public healthcare system.

helpful since the same physician may possess multiple values of other identifiers, which may also not be unique to a single physician.[32]

The next data sources come from the Medical Council and related bodies.

CFM (*Conselho Federal de Medicina*): aggregates information from all Regional Medical Councils in the country. Physicians are required to register with the Regional Medical Council in order to be able to legally offer medical services in the given state. In addition to physicians' CRM (license issued by Regional Medical Councils), the registry contains information on physicians' name, gender, date of birth, university of graduation, and graduation date. The database includes both active and inactive license numbers (i.e., includes all physicians who ever registered in any of the country's Regional Medical Councils).

CNRM (*Conselho Nacional de Residencia Médica*): registry comprising the list of all physicians who ever completed a residency degree in any medical field. Available information includes physicians' name, name of completed residency program, name of medical school, beginning and conclusion dates of training, as well as their CRM. The medical schools' name does not directly map into health facility codes, and residency programs are informed in a free text field (it does not directly map the official medical specialties). Therefore, we only use this source of data to learn whether a physician has completed at least one residency degree by the time they offer medical services.

3.3.2 Data linkage

When linking the two modules (RD and SP) of the SIH/SUS database, we keep information on physicians' CNS id and registered occupation of physicians involved in the provision of the hospitalization's main procedure.

Next, we merge information from CNES/PF on additional physician identifiers and type of employment attachment at the physician-hospital-month level, using CNS number as the key common variable. The additional physician ids comprise the unique identifier

---

[32] Names are long string variables which could easily be misspelled, or changed over time (e.g., change of surname after marriage/divorce or change of first name after official request). CNS is an identifier introduced in the early 2010s by the Ministry of Health, which in the first years of implementation could have been generated multiple times in different hospitals for the same individual provider. CRM is issued at the state level and, therefore, is only unique if used together with the state code where it was issued. Unfortunately, CNES/PF does not inform the state to which the CRM refers to. Although physicians are expected to register the CRM for the given state where the hospital is located, this might not always be the case (e.g., migrant physicians who register right after they move and never update their records, physicians reporting a single CRM in all states where they are active).

(previously generated based on the encrypted CPF), CRM (medical council registration), and physicians' full name.

Finally, we add physician-level characteristics from CFM and CNRM. These include gender, date of birth, university of graduation and date of graduation as well as an indicator of whether the physician has ever completed a residency degree and the respective date when this happened. Given that we have information on exact dates of admission, we are able to compute physicians' age at the time of hospitalization and whether physicians had completed residency by then. To get a measure of university quality, we add information on the university's score (which ranges from 0 to 5) obtained in the National Exam of Students' Performance (ENADE) for the Medicine undergraduate degree.[33]

3.3.2.1 Challenges in data linkage

Our data contains information on hospitalization claims and all physicians involved in each hospital episode. The first challenge we encounter is that the same hospital admission may have been broken into different hospitalization claims. If this is the case, we need to aggregate multiple claims to get unique hospitalization identifiers. The second challenge is that physician identifiers change across different sources of data, and some are more reliable than others in terms of duplicates. Linking hospital activity data (from the Ministry of Health) to physician-level characteristics (from Medical Council) demands some data manipulation. Below, we detail how we address these two challenges.

<u>Aggregating hospital claims into unique hospitalization identifiers</u>

Because the system where the hospitalization claims are entered only allows for a single main procedure per claim, providers are likely to generate multiple claims for the same patient whenever more than one main procedure is performed for the given hospital stay (e.g., childbirth followed by surgery due to complications during delivery). Another restriction imposed by the system is on the maximum number of days of hospitalization. If the spell of a given admission surpasses the threshold for the performed procedure, providers usually "close" the existing claim and generate a new one with the same procedure code and patient details but subsequent dates. Furthermore, when providing long-term care, hospitals

[33] ENADE is used by the Ministry of Education to evaluate higher education degrees every 3 years. We average the scores across the years of 2007, 2010, 2013, 2016, and 2019. Although the exam is not compulsory for some universities, all of them have opted in except USP (University of Sao Paulo). Given that USP is believed to offer the most prestigious Medicine degree in the country (it was ranked first in Latin America in the 2021 Times Higher Education World University), we attributed to it the highest ENADE score of 5 points.

are asked to submit new claims for the same patient every month so that they can be reimbursed before the patient is discharged.[34]

Although we cannot deterministically link hospitalization claims corresponding to the same patient stay given that patient ids are not observed, the richness of our data allows us to aggregate them in quite a reliable way. We use the following variables to match hospitalization claims concerning the same patient stay: start date, end date, granular patient characteristics, and reason for "closure" of hospital claim. When a claim record is closed for the reasons mentioned above, the last variable informs "administrative reasons" (instead of patient discharge or death), and its end date refers to the calendar date when it was closed. Because the system only allows the same patient to be registered in a single hospitalization claim at any given point in time, the original claim needs to be closed before a new one can be generated. Thus, the start date of the new claim should always correspond to the same (or following) day as the end date of the "closed" claim. We use an algorithm that matches claims ending at date t due to "administrative reasons" with claims starting on t or t+1 in the same hospital for patients with same fixed characteristics (gender, date of birth, postcode). We manage to reduce the proportion of records informing "administrative reasons" as reason for closure from 5.78% to 0.69%.[35]

Linking physician-level variables to unique physician identifiers

The merges between hospital activity data and the sources of data containing physician-level characteristics are particularly challenging because we lack a single unique identifier variable available in all the different sources of data.

The CRM number is the most widely used physician identifier in the country as it constitutes their legal license to offer medical services in the given state. However, there are a few limitations that prevent us from using it as the single main individual physician identifier. First, this number is generated at the state level. Because different states might generate the same number, this identifier is only unique if used together with the corresponding state code (e.g., a physician whose license number to operate in Rio de Janeiro is 1234 would use RJ-1234 as her CRM number). While the CFM database informs CRM

---

[34] Another reason why the same patient episode may correspond to multiple claim records is when the patient is transferred across hospitals. In this case, we keep the claims as separate ones given that patients are seen by different physicians after they are transferred between hospitals.

[35] Records tagged as closed due to "administrative reasons" do not include the last claim for the given hospital stay (the last claim's reason for closure is either patient discharge or death); therefore, the share of all claim records corresponding to hospital stays broken into multiple claims in the original data is higher than 5.78%.

number coupled with the state where it was issued, remaining sources (CNES/PF, CNRM) inform CRM without any reference of the state where it was issued. Second, although the CRM code is unique to a single physician (if provided with the state code), the same physician could have multiple CRM codes. This would happen whenever physicians have ever registered in more than one state, or if they re-issued a license within the same state (for instance, after not being active for a period). The CFM database contains all ever-issued state licenses. In each CRM registration, the physician informs all their fixed characteristics (full name, date of birth, university of graduation, and date of graduation degree). We create a homonyms indicator that tags physicians' names associated to different dates of birth.

Below, we explain the process by which we conducted this linkage, comprising multiple steps aimed at reliably maximizing the match rate. First, we add together the variables available in CNRM and CFM. Next, we merge these variables with the set of active physicians in SUS, available at CNES/PF.

To aggregate the variables informed at CNRM and CFM, we use as key variables the CRM and the exact physician's name, which are the two identifiers available in both sources of data.[36] In this first match, we manage to match 90% of physicians who ever reported a residency degree to a corresponding row in the CFM data. In the next step, we match the remaining unmatched rows using as key variables the CRM number and closest physician' name based on an algorithm that computes similarity scores.[37] Finally, for physicians' names tagged as non-homonymous, we use the exact name as a key variable. Less than 4% of CNRM rows (representing all residency diplomas issued in the country) remained unmatched, indicating physicians not found within the Federal Medical Council data.

Lastly, we add information on all physician-level characteristics (from CFM and CNFM) to the set of active physicians in SUS, available at CNES/PF. As explained in Section

[36] Although we have information on state where residency degree was obtained, we do not use it in our merge given that this may not correspond to the state where the CRM was issued. Physicians might, for instance, have an active CRM license in a given state (e.g., where they studied their undergraduate course) and then move to other regions for their residency degree.

[37] As mentioned above, the same CRM may be associated with different physicians in the CFM data if we don't condition on the state where it was issued. Besides, physicians listed in the CNRM data source (i.e., those having completed a residency degree) should be found in the CFM data, given that the latter source contains all licensed providers. In the absence of missing information, physicians would not have been matched in the previous merge because of differences in spelling of reported physician name. To find the closest name (for a given CRM) across the two data sources, we use the algorithm made available through the -stringsim- R command (available in -stringdist- package), which computes similarity scores between strings. We match a row of CNRM to another in CFM that informs the same CRM and physician name with the highest similarity score relative to the former as long as it is above 80%.

3.3.1, the CNES/PF database enables us to generate unique physician identifiers. We restrict this data source to the sample of physicians who ever show up in our hospital claims. For each unique physician identifier, we observe all names and CRMs (without state code) ever reported across different hospital-months where the physician was active. To maximise the match rate, we cross all ever-reported names and CRMs for a given (unique) physician to get all possible combinations of names and CRMs for the same individual physician. We then use these two variables (name, CRM) together with the state code of the hospital (where the physician registered the given CRM) to conduct the merge. We were able to match information for 99% unique physician identifiers from the universe of physicians active in SUS during the analysis' period. Next, we proceed with a second merge using solely name and CRM (no state code) as key variables. For instances of non-homonymous physicians' names, we perform a final merge where we use the full name of the physician as the common variable between the datasets being linked. Only 0.7% of physicians with active records in SUS remain unmatched to the Medical Council's available information on individual characteristics.

3.3.3 Identification of medical specialty

Physicians inform their registered occupation when submitting a hospitalization claim in SIH/SUS. These occupations are mapped into the corresponding medical specialties officially recognised by the Federal Medical Council.

Although physicians inform occupations for each service they provide, we fix specialties at the physician level. By fixing physicians' choice set, we address endogeneity that may arise in physicians' decision on which specialty to report whenever they have multiple options to choose from.[38] We use, therefore, information of *ever* reported occupations to construct the set of all medical specialties of each individual physician.

[38] The specialty declared by the physician must align with their accredited occupations for the specified hospital-month, and it must also be among the eligible provider occupations permitted to perform the given procedure. The system automatically rejects claims if the reported provider occupation is inconsistent in either of these dimensions.

3.3.4 Initial sample of physicians

We observe the activity of 215,842 physicians within SUS during the period between July/2012 and December/2019. This corresponds to more than half of the universe of all physicians in the country, which according to Scheffer et al. (2018) was 414,831 in 2017.[39]

As explained above, we use reported physicians' occupations to construct physicians' specialty sets. The mapping between occupation and official medical specialty is very straightforward, except for two cases. Some hospitals (usually the most prestigious ones) include "resident" as a separate code in the occupation variable. This is replaced by the physician's respective medical specialty once they complete their residency degree. Because we are not able to attribute any medical specialty to physicians who we solely observe practicing under the "resident" occupation, we drop these from our sample (12,195 physicians, or 5.6% of all physicians). We keep, however, those who ever register another occupation. We also drop other 206 physicians (corresponding to 0.1% of the remaining sample), who reported having no medical specialty.

Finally, we also remove from our sample physicians who solely report working as anaesthesiologists. When working as anaesthesiologists, physicians are usually restricted to administering anaesthetics (which is generally already included in the procedure cost) and don't influence the choice of the patient's overall medical treatment. Those who report another occupation (e.g., intensivists), in addition to working as an anaesthesiologist, are kept in our sample.

These initial restrictions narrow down the number of physicians to 189,086. From this universe, our estimation will exploit information on 152,792 physicians (including peers) to estimate peer effects on the cost of hospitalization conducted by a total of 70,770 focal physicians. We will elaborate on how these figures are derived in the concluding part of the next section, after detailing our empirical strategy.

3.4 Methods

3.4.1 Peer definition

Peers are defined as the group of physicians working together in the same hospital at the same month who share at least one medical specialty. Peers at a given point in time are,

[39] This figure includes physicians who solely serve the private sector, i.e., those for whom we do not observe any activity among publicly funded hospitalization claims.

thus, current colleagues who have *ever* acted in the same medical field - and therefore present some similarity in terms of basic medical skills and type of expected treatment decisions.[40] Physicians who take similar decisions as regards to which medical field to specialise are more likely to engage in social interactions in the workplace not only because they tend to have similar medical interests and career aspirations, but also because they are likely to share intrinsic traits and preferences.

In this chapter, we study how physicians' decisions about how to treat their own patients are affected by how their fellows treat their respective patients.

3.4.2 Peer effects and identification challenge

We are interested in understanding how physicians respond to their colleagues when caring for patients in the hospital setting. There are two ways through which individuals are affected by each other (Manski, 1993). The first one is known as *endogenous* peer effects and refers to how individual physicians responds to the behaviour of their peers. The second one, referred to as *contextual* peer effects, relates to physician's responses to the characteristics of their peers. For instance, physicians' decisions may be influenced by the seniority or gender of nearby physicians regardless their behaviour. In this study, we refer to endogenous peer effects simply as peer effects (if not otherwise stated) given that we are mostly interested in effects of this nature.

We are not able to uncover peer effects by simply regressing individual outcome on peer outcome for many reasons. When individual and peer outcomes are both measured at the same time period, inference on peer effects is unfeasible because of simultaneity. This is usually called the reflection problem because one is not able to discern the direction of the relationship between the two outcome variables. One way to address this issue (and make sure the direction under evaluation is from peer outcome to individual outcome) would be to construct the peer outcome variable based on past data. This is also sensible if we are interested in evaluating effects which are not immediate (i.e., not mediated through mimicking peer behaviour). However, even if simultaneity is solved, two other identification challenges remain: endogenous group formation and correlated shocks. First, because individuals are likely to select into the workplace based on their intrinsic preferences. Because

[40] As explained before, we define medical specialties as fixed at the physician level. Consequently, co-workers who have ever provided services within the same medical occupation are considered to be peers, regardless of whether they were colleagues at the time they practiced in these specific medical specialities.

individual choices are largely determined by one's preferences, positively correlated peer behaviour could simply stem from them having positively correlated preferences. Second, because individuals in the same network share the same environment, they are expected to simultaneously respond to the same local shocks.[41] Positively correlated behaviour could, therefore, simply be a consequence of positively correlated shocks. Therefore, physician sorting and common exposure to local shocks would result in outcomes of peers moving together in the same direction, even in the absence of peer effects. To circumvent these challenges and be able to identify how physicians' outcome responds to changes in the outcome of their peers, we need to find econometrically exogenous variation of the latter. We do so by using an instrumental variable research strategy.

3.4.3 Instrumental variable approach

We take advantage of the *not perfectly* overlapping nature of our network to find exclusion restrictions. Because physicians may work simultaneously in multiple hospitals, we observe doctors who work with the peers of the focal physician in hospitals where the focal physician has never worked. In the context of our instrument, we use a restrictive definition of "peers of peers" to include physicians who share a medical specialty and workplace with the peers of the focal physician but not with the focal physician herself. We are able to find non-empty sets of this group because our network of physicians also does not perfectly overlap in terms of medical specialty. Physicians may have multiple medical specialties, and sharing one medical specialty does not mean having identical sets of medical expertise.[42] This makes our results robust to shocks to medical specialty, such as innovations and new scientific knowledge specific to certain areas of care, which are expected to influence physician behaviour as well as peer composition (due to hiring decisions, for example).[43]

---

[41] This second challenge would be an issue even in the context of random allocation of physicians across teams, given that physicians would still be exposed to the same local shocks and supposedly simultaneously respond to them in the same direction (e.g., cutting cost in response to local funding crises), even if their preferences are not correlated.

[42] Many combinations of medical specialties are possible (e.g., general medicine & cardiology; surgery & cardiology; surgery & mastology; obstetrics & mastology). As mentioned in Section 0, many medical specialties (if acquired through medical residency) require a prior specialty. For instance, to start a residency in cardiology, physicians need to have concluded residency in either general medicine or general surgery; to enter residency in mastology, a prior degree in either surgery or obstetrics is needed.

[43] The introduction of a new technology to a medical specialty could lead incumbent physicians to change their behaviour as well as trigger hospitals to hire younger physicians who have more up-to-dated knowledge of frontier technology (and, therefore, are more likely to use it).

Consider a simple example of four physicians working together in a given hospital, who are registered to provide services in the following medical specialties:

- Physician i: clinician and cardiologist
- Physician ii: general surgeon and cardiologist
- Physician iii: cardiologist (only)
- Physician iv: clinician (only)

Physicians i, ii and iii are peers (cardiologists), so are physicians i and iv (clinicians). Peers of peers of physician i will be the set of general surgeons with whom physician ii works in other hospitals. Peers of peers of physician ii will be the set of clinicians with whom physician i works in other hospitals. Peers of peers of physician iii will constitute of both the clinicians who are peers of physician i as well as the general surgeons who are peers of physician ii in hospitals excluded from physician iii's network. Finally, peers of peers of physician iv will be the group of cardiologists with whom physician i practices elsewhere.

The idea behind this choice of instrument is to use exogenous variation in the behaviour of peers that is induced by (exogenous) changes in the composition of their own peers (i.e., their peers' peers) who are not part of the focal physician's network in neither dimension: place of work (i.e., hospital) nor type of work (i.e., medical specialty).

The key necessary assumptions for identification in instrumental variable frameworks are: (1) (enough) correlation between the instrument and the endogenous variable (2) exclusion restriction (i.e., the instrument only affects the outcome of interest through its effects on the endogenous variable), and (3) exchangeability assumption (i.e., absence of unobserved common factors determining both the instrument and the outcome variable). In our context, the first assumption requires characteristics of peers of peers to be good predictors of peer behaviour (we explained above the two channels through which this could happen). The second assumption states that peers of peers' characteristics only affect focal physician behaviour through their impacts on the behaviour of their peers in common. For the third assumption to hold, all factors influencing both the composition of peers of peers and the behaviour of focal physician need to be controlled for. A more rigorous articulation of this requirement is that focal physicians, who are prone to using medical resources in varying extents (i.e., exhibit different potential outcomes), should be associated with similar conditional distributions of peers of peers.

3.4.4 Model specification

Our model specification consists of a linear-in-means model, where we regress the hospitalization cost incurred by focal physicians (i.e., own physician outcome) on the average hospitalization cost of their peers (i.e., average peer outcome) over the past 30 days, along with characteristics of the focal physician and their peers.

Consider the outcome Y_{hf} of a given hospitalization *h* by a focal physician *f*. The average peer outcome, $\overline{Y_{-hf}}$, will be instrumented with the average characteristics of peers of peers, denoted by $\overline{\overline{X_{-hf}}}$. In the end of the section, we explain in detail how these variables are constructed. The model is described below, where the top equation refers to our main regression of interest and the following equation details the first-stage regression.

$$Y_{hf} = \beta\widehat{\overline{Y_{-hf}}} + \delta\overline{X_{-hf}} + \gamma X_{hf} + \eta Pat_h + \pi Hos_h + FE + \nu_{hf} \qquad \textbf{(3.1)}$$

$$\text{where } \overline{Y_{-hf}} \text{ is instrumented with } \overline{\overline{X_{-hf}}}$$

$$\text{and } FE = Diag_h + Mun_h + Mon_h$$

$Diag_h, Mun_h, Mon_h$ represent the set of dummy indicators for diagnosis, municipality, and year/month of hospitalization admission, respectively.[44] Hos_h concerns characteristics of health facility during the month of *h*'s admission. It includes indicators of general *vs* specialised hospitals, teaching *vs* non-teaching hospitals, as well as facilities with general *vs* restricted admission protocols. Patient's demographics (age and gender) are denoted by Pat_h. The vector of focal physician's characteristics, X_{hf}, includes seven variables: gender, age at time of hospital admission, indicator for having graduated from a "top" university,[45] indicators for having completed a residency programme by month of *h*'s admission and having completed residency degree by December/2019 (last month of our

---

[44] Patient diagnosis is reported as of ICD-10. We create a diagnostic indicator variable referring to each one of the 211 official groups across the 22 ICD-10 chapters. This level is detailed enough to inform the reason for hospitalization, but not too detailed to specify the type of treatment. For instance, our indicator variable for childbirth delivery consists of ICD-10 codes between O80-O84 and, thus, does not detail chosen method of childbirth delivery. In case of aggregated hospitalization claims, we keep the diagnosis of the original (i.e., first) claim, as later diagnoses could be endogenous to earlier procedures performed during hospital stay.

[45] We classify as "top" universities those with standardised scores in the National Exam of Students' Performance (ENADE) above the third quartile. The employment attachment variable indicates whether the physician is a staff member (i.e., employee), hired as autonomous physicians, or have other arrangements with the hospital.

analysis' time horizon), indicator of type of employment attachment in hospital where *h* took place during the month of its admission, and an indicator for serving in multiple medical specialties. $\overline{X_{-hf}}$ is a vector of peers' characteristics which includes share of female peers, average age of peers at month of *h*'s admission, share of peers who graduated from top universities, and share of peers who completed residency degree by the month of *h*'s admission. It solely considers peers who were active in the past 30 days (i.e., those conducting hospitalizations considered in $\overline{Y_{-hf}}$). The error terms of our main and first-stage regressions are denoted by ν_{hf} and ε_{hf}. Standard errors are clustered at the municipality level. In robustness checks, we also control for medical specialty fixed effects.

Our main regression of interest exploits exogenous variation of $\overline{Y_{-hf}}$ that arises from variation in $\overline{\overline{X_{-hf}}}$ in the first-stage, conditional on all the model's remaining covariates.[46] Peer effects are measured by β (i.e., endogenous peer effects) and δ (i.e., contextual peer effects). Identification is achieved as long as individual doctors are *at all* affected by their peers, either directly (i.e., contextual effects) or indirectly (i.e., endogenous effects), and such effects don't cancel out.[47]

Summing up, our empirical strategy compares the outcome of physicians of same fixed characteristics within the same municipality who are exposed, in the past 30 days, to peers of same (average) observed characteristics but different (average) behaviour resulting from having been, themselves, exposed to peers of peers of different characteristics. It does so while accounting for heterogeneity in patient observables (demographics and diagnosis), hospital type (in terms of specialised care, teaching status, and referral protocols), and physician-hospital employment attachment (whether officially employed by the hospital or not), in addition of monthly time trends.

Below, we explain how the peer outcome variable and its instrument are constructed.

<u>Constructing peer outcome variable:</u> ($\overline{Y_{-hf}}$) is constructed as the average cost among all hospitalizations which were concluded by peers of physician *f* in the respective hospital

[46] Note that we don't need to control for other factors that predict the average peer outcome variable (e.g., patient casemix, etc) given that we will only exploit exogeneous variation from this variable to study its effect on the outcome of the focal physician.

[47] For more details, see Bramoullé et al. (2009).

during the 30 day-period prior to the focal hospitalization's admission date. We do not include among peer hospitalizations those co-led by the focal physician.

Identifying peers of peers: we identify all physicians who are peers of the focal physician's peers, whose hospitalizations contribute to $\overline{Y_{-hf}}$, while working in other hospitals during the corresponding calendar month at the start of the 30-day window used for constructing $\overline{Y_{-hf}}$.[48] Next, we exclude those whom we ever observe working in the same hospital or sharing a medical specialty with the focal physician. It is important to emphasize that the set of peers of peers associated with a given focal physician's peer varies over time.

Constructing instrument: $\overline{\overline{X_{-hf}}}$ is constructed using a two-step approach. First, we compute the average characteristics of peers of peers directly associated with each peer hospitalization included in $\overline{Y_{-hf}}$. Then, we calculate the mean of these averages across all relevant peer hospitalizations included in the latter variable.[49]

3.4.5 Final estimation sample

Our data is at the hospitalization-physician level. We start from a dataset that includes all hospitalizations conducted by the initial sample of 189,086 physicians described in Section 3.3.4, for which a non-anaesthesiology occupation is reported.[50] When several physicians work in a hospitalization, we use the same hospitalization cost as the outcome variable for all physicians involved. Although the outcome is identical for observations of different physicians performing the same hospitalization, peer outcome is not (peer group depends on the medical specialty of the focal physician in consideration).

---

[48] Say, the focal patient is admitted at day d. Peer hospitalizations will be those concluded during the time interval [d-31, d-1]. Peers of peers will be those with whom these peers have worked elsewhere during the calendar month corresponding to d-31.

[49] This approach maintains consistency with the structure of our model. In our main regression, the outcome of the focal hospitalization is determined by the average characteristics of direct peers. In the first stage regression, this is analogous to modelling the mean peer outcome based on the mean of the average characteristics of the peers of these peers whose outcomes we are trying to explain. The instrument has proven to be stronger when constructed in this manner rather than by taking the simple average among all peers of peers, regardless of how frequently they are associated with the peer outcome in question. This is intuitive given that the former applies higher weight to the peers of peers who contributed to a higher number of peer hospitalizations.

[50] We exclude activity from physicians who report providing anaesthesia services, as anaesthesiologists typically do not influence treatment plans. Nonetheless, our estimation sample includes hospital claims where these physicians reported providing services in any other medical specialty (e.g., intensive care).

We restrict our estimation sample to hospitalizations by focal physicians for which we can observe all recent peer activity to safeguard against mismeasurements in the peer outcome variable. Because we are only able to observe all activity for public hospitals, we only keep focal hospitalizations that took place in these hospitals.[51] Second, we only include observations where the focal physician has exclusively worked in one hospital over the past 30 days and is not affiliated with private hospitals during the current month and the month before.[52] If we were to include physicians working in multiple hospitals, we would need to account for the outcomes of peers they were exposed to in those other hospitals as additional explanatory variables. This would make our model more complicated and the estimation very computationally demanding. It is important to emphasize that these restrictions apply exclusively to focal physicians and do not extend to peers (and peers of peers) considered in the right-hand side of the equation.

Lastly, our estimation sample will include only those observations where both the peer outcome and corresponding instrument are non-missing. To meet this criterion, the focal physician must have been exposed, within the past 30 days, to at least one peer who is associated with peers of peers during the relevant calendar month.

Our estimation sample consists of 13,502,212 observations, at the hospitalization-physician level, which comprise a total of 12,576,646 focal hospitalizations performed by 70,770 physicians.[53] Later in the next section, we show that the distribution of the outcome variable in our final estimation sample is very similar to that of the unrestricted sample (Figure 3.1).

It's worth noting that, in estimating our model, we use data from a substantially larger pool of hospitalizations and physicians. This expanded dataset is necessary to construct explanatory variables for peers' characteristics and hospitalizations, as well as the instrumental variable, which is based on the characteristics of peers of peers. Specifically, we

---

[51] Although we can observe SUS hospitalizations that take place in private hospitals, we have no access to data on privately funded hospitalizations. As a result, we do not observe all activity to which physicians performing SUS hospitalizations are expose to and, therefore, cannot consistently estimate peer effects. Besides, by restricting our estimation sample to hospitalizations in public hospitals, we exploit variation among more homogeneous hospitals.

[52] As explained in Section 3.3.1, the CNES/PF database provides information on all accredited hospitals in the country to which physicians are affiliated in a given month. This allows us to learn whether the focal physicians in our data were affiliated to private hospitals around the time of the focal hospitalization.

[53] The total number of observations exceeds the total number of hospitalizations because some hospitalizations involve multiple physicians. Specifically, 6.5% of hospitalizations in our sample involve more than one focal physician.

use information on 152,792 unique physicians (including focal physicians, peers, and peers of peers). We refer to this latter group, which represents 80% of the total universe of physicians described in Section 3.3.4, as our final sample of physicians.

3.5 Descriptive statistics

Among the focal hospitalizations in our sample, the average duration was 8 days, with an average total cost of R$1,369 (equivalent to approximately US$340 as of December 2019). This cost encompasses various expenses, including fees for medical interventions, ICU utilization, and auxiliary services such as additional diagnostic tests and consultations conducted during the hospital stay. The average age of patients is 39 years, with females constituting 60% of the total. Nearly 90% of hospitalizations conclude with patient discharge, while the remaining cases involve either patient death or transfer to other facilities. Notably, 83% of hospitalizations occur in general hospitals, with 84% taking place in facilities offering general admission processes, including both spontaneous and referred patients. Additionally, 45% of these hospitalizations occur in teaching hospitals.

Table 3.1 presents summary statistics at the observation level. It shows characteristics of the focal hospitalizations, the hospital where they took place, and the focal physicians conducting them. Among all hospitalizations, 38% were conducted by female providers, 31% by physicians having graduated from a top university, and 48% by physicians holding residency degrees at the time of admission. Physicians were, on average, 45 years old at the time of patient admission, and 36% of episodes were led by physicians registered in more than one medical specialty. Roughly 2/3 of hospitalizations were conducted by physicians who were employed as staff members at the respective hospitals.

Among the focal hospitalizations in our sample, the average duration was 8 days, with an average total cost of R$1,369 (equivalent to approximately US$340 as of December 2019). This cost encompasses various expenses, including fees for medical interventions, ICU utilization, and auxiliary services such as additional diagnostic tests and consultations conducted during the hospital stay.[54] The average age of patients is 39 years, with females constituting 60% of the total. Nearly 90% of hospitalizations conclude with patient discharge, while the remaining cases involve either patient death or transfer to other facilities.

[54] While certain basic diagnostic tests, such as blood tests, electrocardiograms, and X-rays, are not charged for under the national tariff reimbursement table, more advanced procedures like ultrasounds, MRIs, and CT scans incur hospitalization-level charges.

Notably, 83% of hospitalizations occur in general hospitals, with 84% taking place in facilities offering general admission processes, including both spontaneous and referred patients. Additionally, 45% of these hospitalizations occur in teaching hospitals.

Table 3.1: Descriptive statistics: focal hospitalizations

	N. unique	Mean	SD
Municipality	1,245		
State capital (%)		48%	-
North (%)		10%	-
Northeast (%)		33%	-
Southeast (%)		41%	-
South (%)		7%	-
Midwest (%)		8%	-
Health facility	1,797		
General hospital (%)		83%	-
Specialised hospital (%)		15%	-
Teaching status (%)		45%	-
General admission (%)		84%	-
Focal hospitalization	12,576,646		
Cost (R$), nominal		1,369	4,790
Cost (R$), real (prices of Dec/2019)		1,633	5,695
Ln cost, nominal		6.41	1.06
Ln cost, real		6.58	1.07
Duration (n days)		8.04	15.96
Surgical procedure (%)		43%	-
ICU use (%)		6%	-
Diagnostic tests (%)		75%	-
Patient gender, female (%)		60%	-
Patient age (years)		39.5	23.9
Final status: discharge (%)		89%	-
Final status: death (%)		6%	-
Final status: transference to another unit (%)		4%	-
Focal physician	70,770		
Formal employment attachment, i.e., staff (%)		73%	-
Autonomous provider (%)		22%	-
Female (%)		38%	-
Age at time of admission		44.7	12.3
University's quality score (0-5)		2.92	0.81
Top university (%)		31%	-
Residency degree at time of admission (%)		48%	-
Residency degree, ever (%)		55%	-
Number of specialties		1.45	0.70
Multiple specialties (%)		36%	-
Number of observations	13,502,212		

Notes: Observations correspond to our estimation sample (focal hospitalization-physician level). The first column presents information on the total number of focal hospitalizations, hospitals, and the respective municipalities where the hospitalization took place, and physicians who conducted them. Standard deviations (SD) are only presented for continuous variables. There are 27 state capitals (i.e., municipalities corresponding to the capital of one of the 27 Brazilian states). Costs are deflated using the Extended National Consumer Price Index, IPCA, at the year/month level.

Table 3.2 presents the distribution of focal hospitalizations across patient diagnoses. Nearly ¼ of these hospitalizations are childbirth-related episodes, followed by medical issues associated with the digestive (12%), circulatory (10%), and respiratory (9%) systems, as well as external causes (11%). Together with genitourinary conditions, infectious diseases, and neoplasms, these categories account for 84% of all focal hospitalizations.

Table 3.2: Distribution of focal hospitalizations by ICD-10 chapters

Diagnosis	ICD-10 Chapter	N.	%
Pregnancy-related episodes	XV	3,295,926	24.4%
Digestive system diseases	XI	1,601,696	11.9%
Injury, poisoning	XIX	1,446,348	10.7%
Circulatory system diseases	IX	1,276,609	9.5%
Respiratory system diseases	X	1,163,160	8.6%
Genitourinary system diseases	XIV	927,092	6.9%
Infectious diseases	I	857,573	6.4%
Neoplasms	II	740,860	5.5%
Musculoskeletal system diseases	XIII	352,592	2.6%
Perinatal period conditions	XVI	305,962	2.3%
Other factors	XXI	277,562	2.1%
Endocrine diseases	IV	276,256	2.1%
Other symptoms	XVIII	242,367	1.8%
Nervous system diseases	VI	194,651	1.4%
Skin diseases	XII	144,724	1.1%
Blood / Immune system diseases	III	139,856	1.0%
Mental disorders	V	104,030	0.8%
Congenital abnormalities	XVII	87,474	0.7%
Eye diseases	VII	48,108	0.4%
Ear diseases	VIII	17,525	0.1%
Other external causes	XX	1,831	0.0%
Codes for special purposes	XXII	10	0.0%
Number of observations		13,502,212	100%

Notes: The table presents, in descending order, the number of focal hospitalizations by diagnosis, as of ICD-10 chapters. In case of aggregated claims (concerning the same hospital stay), we consider the diagnosis of the initial claim.

Table 3.3 displays descriptive statistics of peer hospitalizations, as well as the composition of both peers and peers of peers, identified according to the criteria outlined in Section 3.4.4. Focal physicians were exposed, within the 30-day period prior to the admission of their conducted hospitalizations, to an average of 299 peer hospitalizations, which had an average cost of R$1,264. These hospitalizations were overseen by an average of 31 peers. Finally, we identified an average of 81 physicians who share a workplace and medical specialty with the peers of the focal physician, but neither attribute directly with the focal physician herself (i.e., peers of peers). The table presents average characteristics of these groups. Peers have an average age of 44 years, with 34% being female, 31% having graduated from top universities, and 55% having completed a residency degree. Peers of peers exhibit a lower proportion of females (29%). In addition to presenting information on peers' and peers of peers' characteristics, Table 3.3 also reports the number of distinct identifiers of peer hospitalizations (~25 million), peers (~100 thousand), and peers of peers (~145 thousand) used in our estimation.[55]

Table 3.3: Descriptive statistics: (active) peers and associated peers of peers

	N. unique	Mean	SD
Peer hospitalizations	24,722,619		
Count per focal obs		299	258
Avg outcome			
Cost (R$), nominal		1,264	1,408
Ln cost, nominal		6.38	0.48
Peers	101,769		
Count per focal obs		31	30
Avg characteristics			
Female (%)		34%	25%
Age in given calendar month (years)		44.0	6.1
Top university (%)		31%	27%
Residency degree in given calendar month (%)		55%	24%
Peers of Peers	144,087		
Count per focal obs		81	115
Avg characteristics			
Female (%)		29%	23%
Age in given calendar month (years)		44.3	6.1
Top university (%)		32%	28%
Residency degree in given calendar month (%)		57%	25%
Number of observations	13,502,212		

Notes: Observations correspond to our estimation sample (focal hospitalization-physician level). The first column presents information on the total number of unique peer hospitalizations, peers (who performed such hospitalizations), and the associated peers of peers used in our estimation. Our main regressor of interest, peer outcome, is averaged among hospitalizations conducted by peers in the same hospital which were concluded in the 30 days prior to the admission of the focal hospitalization. We report basic statistics on the total number of peer hospitalizations used to compute this average, as well as the number of different peers who conducted them. Additionally, the total number of peers of peers whose average characteristics we use as instruments are also presented in the table. Average characteristics are presented for both peers and peers of peers.

For the total of 152,792 unique physicians contemplated in our estimation, Table 3.4 outlines the distribution of medical specialties for these physicians, listed in descending order. General medicine stands out as the primary specialty, with half of the physicians practicing in this field (i.e., referred to as clinicians). Clinicians provide general medical care, most of which is covered during undergraduate medical training, although some may pursue further specialization in the field. These programs are highly sought-after, providing valuable clinical experience and serving as a prerequisite for entry into more specialized residency programs.

[55] It is worth noting that our peer (average) outcome variable, comprising of all peer hospitalizations in the past 30 days, is instrumented with the characteristics of peers of peers derived from only a subset of such hospitalizations (i.e., not all peer hospitalizations are associated with peers of peers). The strength of the instrument will reflect the extent to which it explains the average outcome among all peer hospitalizations, including those without any associated peers of peers. Figure B.1 illustrates the distribution of the number of peer hospitalizations, number of peers, and number of peers of peers used to construct the corresponding summary measures. Additionally, it depicts the distribution of the proportion of peer hospitalizations whose peers are associated with peers of peers.

The last column in the table reveals that nearly half of physicians specializing in general medicine hold a residency degree. The next four specialties also offer direct access to residency programs: general surgery (20%), paediatrics (14%), obstetrics/gynaecology (14%), and orthopaedics (8%). Other specialties are associated with less than 5% of physicians (those with less than 1% are excluded from the table). The three columns to the left present summary statistics on gender, age, and university quality of physicians registered in each medical specialty.

The fourth column of Table 3.4 indicates the proportion of physicians registered in the given medical specialty who also provide services in another medical field. Approximately 53% of clinicians and 77% of general surgeons also act in another medical specialty. In highly specialized surgical areas (e.g., surgical oncology), almost all physicians provide services in another medical specialty, typically general surgery. Among those least likely to specialize in another field are paediatricians, obstetricians, and orthopaedists.[56] Table 3.5 replicates this table for the sample of focal physicians. The percentage of focal physicians reporting more than one specialty is 30.8%, while the analogous proportion considering all physicians in our estimation is 36.2%. These statistics, along with other average characteristics, such as age, the percentage with a residency degree, and the percentage of graduates from top universities, can be found in the last row of the respective tables.

[56] We extract medical specialty information from the occupations that physicians report being active in, as indicated when submitting hospitalization claims. In other words, we only consider medical specialties which physicians report claims for. The table indicates that 100% of physicians registered in anaesthesiology are also listed in another specialty because we restricted the data to exclude physicians working exclusively as anaesthesiologists.

Table 3.4: Physician-level distribution of medical specialties: *all physicians*

Medical specialty	N.	%	% multiple specialties	% females	Avg. age	% top uni.	% residency
General medicine (clinicians)	76,029	49.8%	53%	35%	44	30%	48%
General surgery	30,263	19.8%	77%	19%	47	36%	66%
Paediatrics	21,771	14.2%	31%	71%	47	31%	61%
Obstetrics/Gynaecology	21,099	13.8%	42%	53%	48	32%	62%
Orthopaedics	12,062	7.9%	35%	6%	45	35%	64%
Cardiology	7,456	4.9%	62%	26%	49	34%	47%
Intensive care medicine	5,085	3.3%	73%	41%	46	33%	59%
Anaesthesiology	4,952	3.2%	100%	19%	48	36%	56%
Urology	3,692	2.4%	71%	3%	48	42%	75%
Vascular surgery	3,647	2.4%	78%	21%	46	38%	71%
Clinic oncology	3,477	2.3%	75%	44%	46	41%	77%
Neurology	3,337	2.2%	64%	29%	47	41%	63%
Plastic surgery	2,982	2.0%	64%	23%	48	44%	70%
Neurosurgery	2,863	1.9%	63%	11%	47	41%	60%
Cardiovascular surgery	2,560	1.7%	84%	13%	50	42%	55%
Nephrology	2,508	1.6%	64%	47%	47	38%	68%
Otorhinolaryngology	2,122	1.4%	43%	37%	44	46%	59%
Ophthalmology	2,097	1.4%	23%	42%	42	44%	55%
Surgical oncology	2,045	1.3%	98%	17%	47	39%	77%
Gastroenterology	1,908	1.2%	75%	35%	50	39%	64%
Infectiology	1,623	1.1%	56%	54%	47	38%	79%
Gastrointestinal surgery	1,534	1.0%	93%	11%	50	43%	67%
Paediatric surgery	1,481	1.0%	69%	41%	51	39%	68%
All physicians	**152,792**	**100%**	**36%**	**39%**	**45**	**33%**	**57%**

Notes: The table considers all physicians for whom information was used in the estimation, including focal physicians, peers, or peers of peers. Medical specialties refer to those officially recognised by the Federal Medical Council. Physicians serving in the general medicine specialty are referred to as clinicians. The second and third columns present, respectively, the number and share of physicians in the given specialty. Because the same physician is considered across different medical specialties, shares sum to more than 100%. The fourth column informs the share of physicians in the given medical specialty who are also registered in another specialty. This column informs 100% of physicians registered in anaesthesiology as a direct consequence of how we restricted the data to exclude those working exclusively as anaesthesiologists. The last four columns present average characteristics of all physicians registered in the given medical specialty. Age and residency degree status are computed as of December/2019. We only show medical specialties with at least 1% registered physicians.

Table 3.5: Physician-level distribution of medical specialties: *focal physicians*

Medical specialty	N.	%	% multiple specialties	% females	Avg. age	% top uni.	% residency
General medicine (clinicians)	34,329	48.5%	45%	39%	42	27%	48%
General surgery	12,570	17.8%	70%	21%	44	33%	74%
Paediatrics	11,502	16.3%	29%	75%	47	29%	63%
Obstetrics/Gynaecology	10,705	15.1%	33%	59%	46	30%	67%
Orthopaedics	5,594	7.9%	28%	6%	43	32%	69%
Cardiology	2,039	2.9%	61%	31%	47	29%	54%
Intensive care medicine	2,098	3.0%	84%	45%	44	31%	60%
Anaesthesiology	1,420	2.0%	100%	23%	45	35%	61%
Urology	1,492	2.1%	72%	4%	44	40%	85%
Vascular surgery	1,522	2.2%	74%	27%	43	36%	80%
Clinic oncology	1,062	1.5%	73%	52%	45	36%	82%
Neurology	1,152	1.6%	62%	35%	45	36%	68%
Plastic surgery	1,361	1.9%	63%	24%	46	43%	76%
Neurosurgery	1,184	1.7%	56%	14%	44	37%	68%
Cardiovascular surgery	647	0.9%	83%	19%	47	42%	67%
Nephrology	822	1.2%	63%	55%	45	34%	75%
Otorhinolaryngology	640	0.9%	37%	43%	43	45%	65%
Ophthalmology	422	0.6%	36%	41%	43	42%	58%
Surgical oncology	712	1.0%	98%	19%	46	36%	79%
Gastroenterology	558	0.8%	75%	49%	46	34%	73%
Infectiology	831	1.2%	59%	54%	47	35%	80%
Gastrointestinal surgery	583	0.8%	96%	14%	47	39%	78%
Paediatric surgery	731	1.0%	65%	48%	49	37%	76%
All physicians	**70,770**	**100.0%**	**31%**	**43%**	**44**	**30%**	**60%**

Notes: Table 3.4 for the sample of focal physicians (i.e., those conducting focal hospitalizations in our estimation sample). We keep the same order and list of specialty of Table 3.4 for ease of comparison. Medical specialties refer to those officially recognised by the Federal Medical Council. Physicians serving in the general medicine specialty are referred to as clinicians. The second and third columns present, respectively, the number and share of physicians in the given specialty. Because the same physician is considered across different medical specialties, shares sum to more than 100%. The fourth column informs the share of physicians in the given medical specialty who are also registered in another specialty. This column informs 100% of physicians registered in anaesthesiology as a direct consequence of how we restricted the data to exclude those working exclusively as anaesthesiologists. The last four columns present average characteristics of all physicians registered in the given medical specialty. Age and residency degree status are computed as of December/2019.

Turning back to the observation level of our analysis, Table 3.6 provides information on the main specialty of the physician overseeing the focal hospitalization (i.e., the most frequently reported), the specialty most commonly shared with their peers, and the specialty most frequently shared between peers and peers of peers. While the first two specialties are often correlated (coinciding for 81% of observations), the third is always different from the first two by design, as we select peers of peers who do not share any specialty with the focal physician.

The second column of the table reveals that nearly 1/3 of observations in our sample are conducted by physicians who frequently work as clinicians, 1/4 by gynaecologists, and 1/7 by general surgeons. General medicine is the most commonly shared specialty between focal physicians and their peers, as indicated by 45% of observations in our estimation sample. The last column indicates that general medicine and general surgery are the most frequently shared specialties between peers and peers of peers.

Table 3.6: Distribution of medical specialties at the observation level

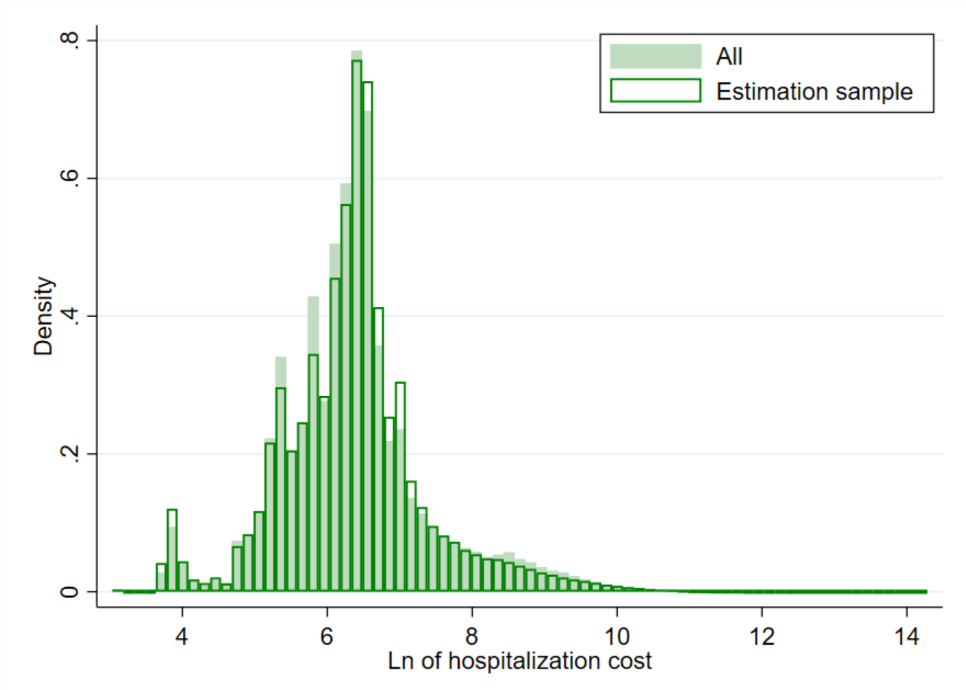
Medical specialty	Most freq. reported by Focal physician	Shared Focal-Peers	Shared Peers-Peers of Peers
General medicine (clinicians)	31.51	44.78	33.84
Obstetrics/Gynaecology	25.83	21.40	9.04
General surgery	14.19	13.61	35.71
Orthopaedics	8.06	5.74	5.78
Paediatrics	7.50	6.74	5.37
Cardiology	1.46	1.20	1.41
Urology	1.34	0.52	0.55
Vascular surgery	1.29	0.52	0.46
Plastic surgery	1.12	0.55	0.36

Notes: Medical specialties refer to those officially recognised by the Federal Medical Council. Physicians serving in the general medicine specialty are referred to as clinicians. A given physician can be registered in multiple medical specialties. All figures as in %, representing the proportion of observations in our estimation sample (consisting of 13,502,212 observations, at the hospitalization-physician level). For each observation, we constructed the following three specialty variables: (i) the medical specialty most frequently reported by the focal physician (across all their submitted claims), (ii) the specialty of the focal physician shared with the largest number of peers who were active in the past 30 days (i.e., considered in the peer outcome variable), and (iii) the specialty of active peers shared with the largest number of peers of peers. The distribution of each one of these are presented in the three last columns of the table. We only show medical specialties representing at least 1% of observations in the second column.

Finally, we present the distribution of our outcome variable of interest: the natural logarithm of the total cost of hospitalizations. This log transformation is commonly adopted for variables with highly skewed distributions, which is typically the case with costs, to render the distribution more closely resembling a "bell curve".[57] The histogram below demonstrates that this is reasonably achieved. Additionally, it illustrates that the distribution of focal hospitalizations' cost closely mirrors that of the entire universe of hospitalizations conducted during our analysis period. Based on our model specification, we do not need to adjust for inflation in our cost measures (i.e., because our outcome variable is specified as log, time fixed effects would capture inflation over time).

[57] Taking the log of a variable is problematic in case of high frequency of zeros (given that the logarithm function is undefined at this value), which is not our case – zero cost has only been reported for 0.01% of hospitalizations.

Figure 3.1: Distribution of hospitalizations' (ln) cost

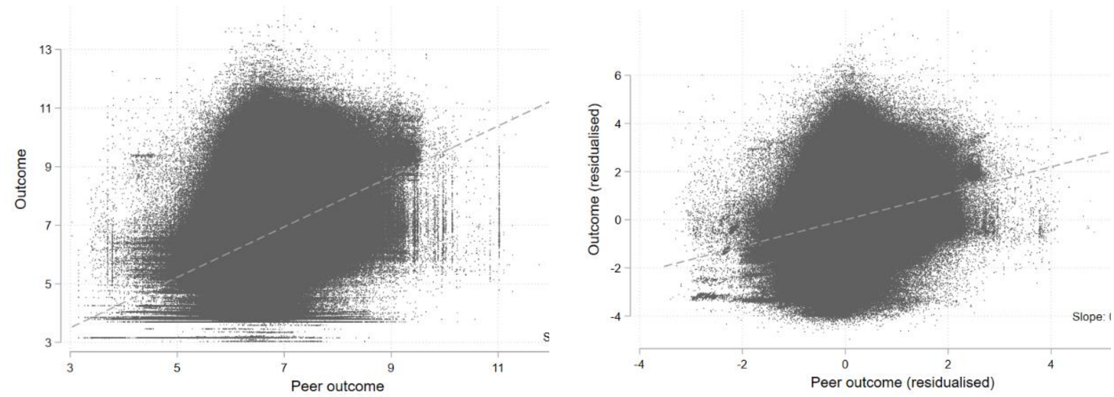


Notes: Costs are measured in R$, in nominal terms. The distribution in light green considers all hospitalizations from August 2012 to December 2019 (a total of 79,977,306) conducted by the universe of 189,086 physicians described in Section 3.3.4. The distribution outlined in darker green considers the focal hospitalizations in our estimation sample: a total of 12,576,646 focal hospitalizations, conducted by 70,770 focal physicians during this time period. This includes hospitalizations in public hospitals by physicians who only worked in the given hospital during the past 30 days and are not affiliated to private hospitals during the current and previous months, as described in Section 3.4.5. The time period of our analysis elapses from July/2012 until December/2019.

The subsequent plots demonstrate that, as anticipated, the outcome of the focal physician exhibits a positive correlation with the average outcome of their peers. Panel B shows that the slope becomes less steep once we account for the model's control variables, yet it remains above 0.5. This cannot, however, be interpreted as evidence of peer effects, as the positive relationship could simply be driven by common unobserved characteristics associated with physicians' preferences or by their reactions to local shocks to the hospitals where they work or the medical specialties in which they practice, as previously discussed. In the next section, we will investigate the first stage variation which will be exploited to causally identify peer effects.

Figure 3.2: Relationship between focal outcome and peer outcome

Panel A: Raw data Panel B: Residualised



Notes: Each dot represents one observation in our estimation sample. Outcome is measured as the natural logarithm of hospitalization cost. The y-axis refers to focal physician's outcome at the hospitalization level (i.e., ln cost of the focal hospitalization), whereas the x-axis concerns the average ln cost of all hospitalizations concluded by peers of the focal physician during the 30-day window prior to the admission of the focal hospitalization. Panel A plots the raw data, while Planel B adjusts both outcomes for the model's covariates and fixed effects (detailed in Section 3.4.4).

3.6 Selection of instrument and model specification

Our first-stage results stem from comparing the average cost of observably similar groups of physicians (i.e., peers, in our model) who were recently exposed, when working in other hospitals, to different groups of colleagues in shared medical specialties (i.e., peers of peers).

Table 3.7 shows the main first-stage coefficients of just-identified regressions where we use each one of the available characteristics as a single linear instrument. Among all IV candidates, gender of peers of peers is the only variable that helps explain peer outcome – the coefficients of all the remaining characteristics are not statistically different from zero. The first-stage "Effective" F-statistic is around 20 when using the share of female peers of peers as the sole instrument, while this statistic is around 1 when considering the other characteristics as the excluded instrument.[58] When looking at the predictive power of peers' average characteristics in explaining their *own* average hospitalization cost, gender and residency degree are the only relevant ones. Surprisingly, peers' age and the quality of their university of graduation do not help predict costs. Although peers' costs are explained by their own residency degree status, their exposure to higher or lower proportions of co-workers (i.e., peers of peers) who have completed residency programmes does not increase

[58] To detect weak instruments, we employed the "Effective" F-statistic, as recommended by Olea & Pflueger (2013), which is widely regarded in the literature as the most suitable statistic when the estimation involves clustered standard errors (Andrews & Stock, 2018)

the explanatory power. With these results in mind, we proceed with peers of peers' gender as a singular instrument.

Table 3.7: First-stage results of just-identified regressions: linear specification

Outcome:	Average of Ln of Peers' Hospitalization Cost			
IV:	Gender	Age	Top uni.	Residency
Peers of Peers' characteristics (i.e., IV)				
% female	-0.093***			
	(0.020)			
average age		-0.001		
		(0.001)		
% top university degree			-0.018	
			(0.023)	
% residency degree				-0.026
				(0.021)
Peers' characteristics				
% female	-0.344***	-0.354***	-0.355***	-0.355***
	(0.066)	(0.068)	(0.068)	(0.068)
average age	0.002	0.002	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
% top university degree	-0.047	-0.045	-0.044	-0.043
	(0.044)	(0.044)	(0.043)	(0.043)
% residency degree	0.227***	0.227***	0.227***	0.225***
	(0.050)	(0.050)	(0.050)	(0.049)
Diagnostic group FE	X	X	X	X
Municipality FE	X	X	X	X
Year/month FE	X	X	X	X
Effective F-stat	20.08	0.84	0.77	1.65
Observations	13,502,212	13,500,271	13,497,451	13,498,708

Notes: All remaining regressors described in Section 3.4.4 (focal physician's characteristics, patient demographics, and health facility observables) are omitted from the table. Standard errors are clustered at the municipality level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

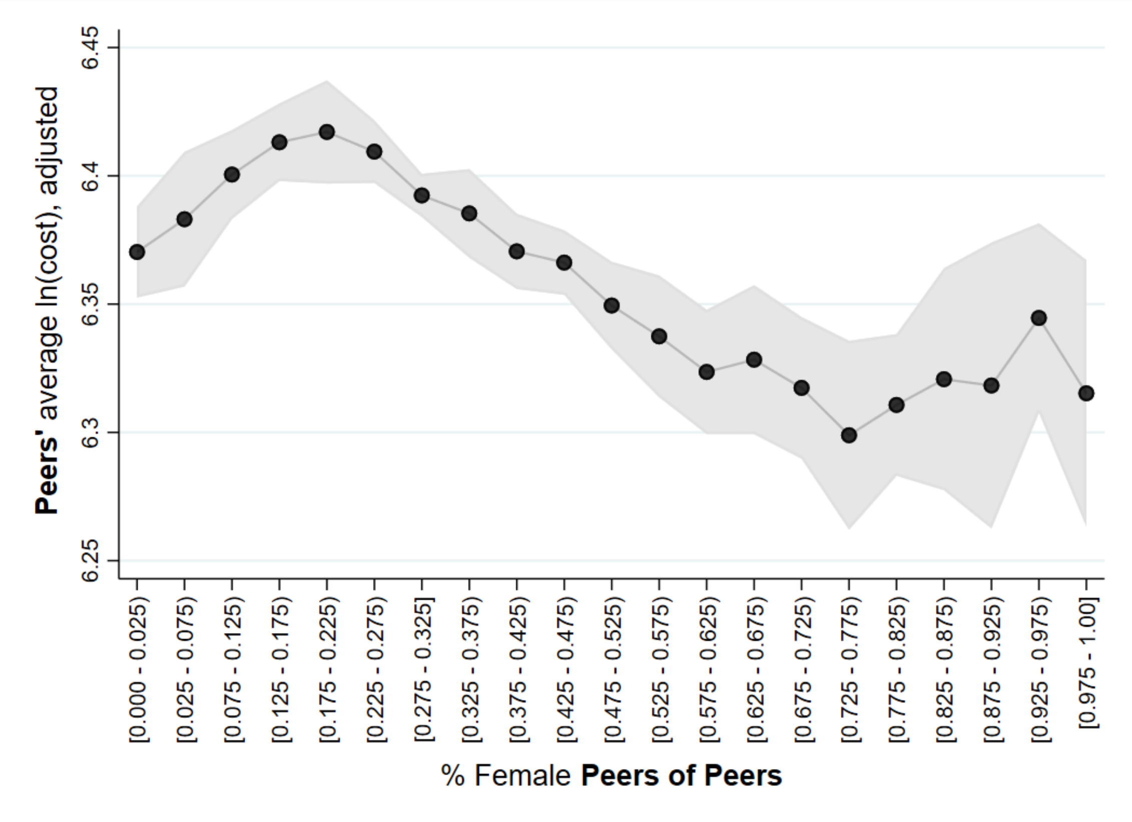
The previous table points to a linear negative relationship between female gender of peers of peers and peers' average hospitalization cost. Next, we investigate this first-stage relationship more closely, without imposing any parametric assumption. We ran the same regression as above but, instead of adding the proportion of females linearly (for both peers and peers of peers), we replaced it by a set of indicator variables of small interval brackets. Figure 3.3 plots the adjusted mean of the peer outcome variable (in the y-axis) for each one of these brackets (in the x-axis). This adjusted mean represents the predicted average costs of peers (measured in natural logarithm) for different levels of the proportion of females among peers of peers, with the contribution of other covariates held constant at their mean values.[59]

While Figure 3.3 (Panel A) shows a predominantly negative relationship between peer cost and the share of female peers of peers, an interesting pattern emerges. The negative association is primarily observed within the middle of the distribution, where the share of females among peers of peers ranges from 25% to 75%. However, within the range of 0% to 20%, precise estimates unveil a positive relationship between peer cost and the proportion of females among peers of peers. The confidence interval widens notably beyond 75%, reflecting fewer observations where females constitute more than 3/4 of peers of peers, as depicted in Panel B.

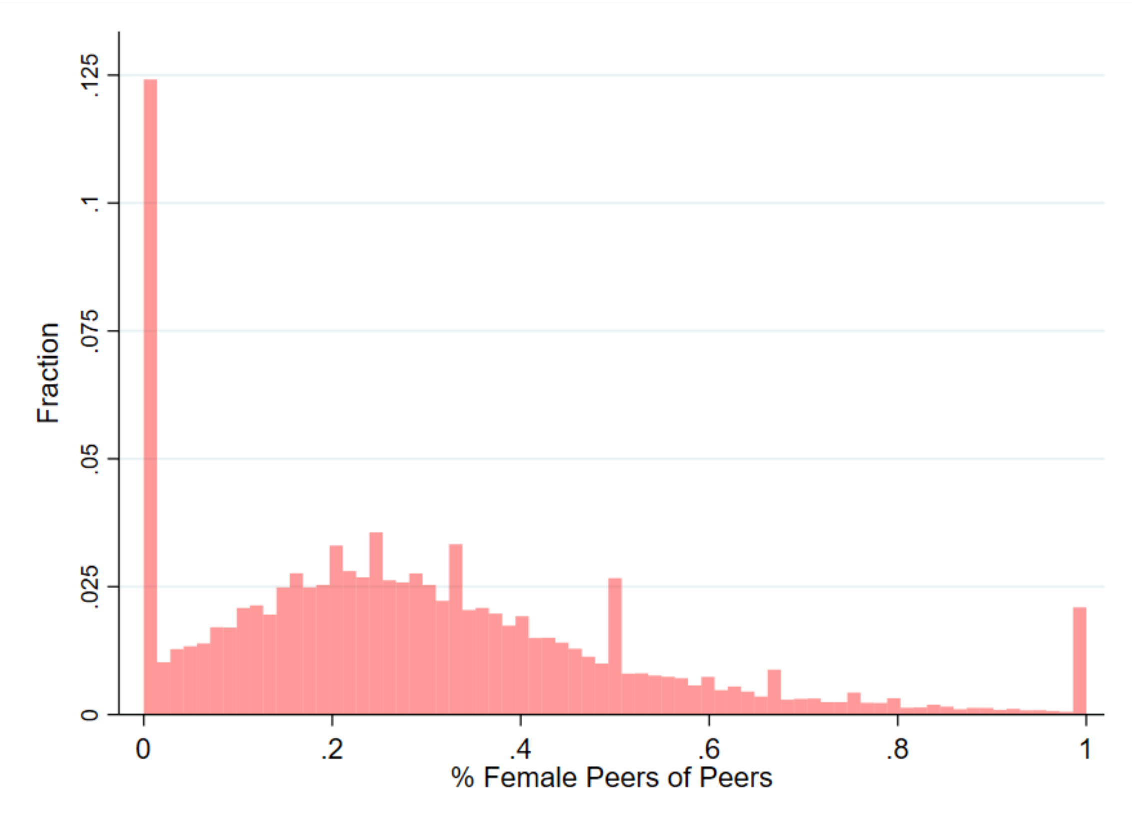
[59] The adjusted mean is computed as the sum between the coefficient estimate of the dummy variable corresponding to the given female shares (informed in the x-axis) and the average contributions of all other regressors. These contributions are computed as the mean product between the value of the covariate and their coefficient estimate.

Figure 3.3: Relationship between (adjusted) peer outcome and IV

Panel A: Average of Peer Outcome by IV interval, adjusted by covariates



Panel B: IV distribution



Notes: Panel A shows the relationship between our endogenous variable (i.e., peer outcome) and its instrument (i.e., share of female physicians among peers of peers), while controlling for all covariates and fixed effects included in Equation (3.1) in Section 3.4.4. To obtain this non-parametric relationship, we estimate a version of our first-stage regression where we model female shares as indicator variables for small interval brackets while keeping other covariates as specified in our baseline model. The plot shows the adjusted mean of peer outcomes (in the y-axis) across each one of these brackets (in the x-axis), while fixing the contribution from other model's covariables at their average value (i.e., mean among the product between the estimated coefficient and the observed values). The 95% confidence interval is represented in light grey. Panel B plots the distribution of the proportion of females among peers of peers across our estimation sample, which we use to instrument our peer outcome variable.

To assure ourselves that the association depicted by Figure 3.3 is not driven by differences in cost and gender composition across medical specialties, we add specialty fixed effect (FE) to the regression.[60] Considering that physicians may hold multiple medical specialties, focal physicians might have peers in different specialties. Peers could also be affiliated with numerous other medical specialties besides those shared with the focal physician. For these reasons, there are a few alternative ways to condition results on medical specialty.[61] The first approach is to include fixed effects (FE) representing the most frequently reported specialty by the focal physician.[62] Another approach is to include FE representing the specialty shared with the largest number of peers.[63] A third option involves a more saturated model, wherein we augment the model with FE representing not only the focal physician's specialty shared with the largest number of peers but also the peers' specialty shared with the largest number of peers of peers.

Figure 3.4 illustrates the same plot while controlling for specialty FE across the three specifications detailed in the previous paragraph. These statistically similar patterns confirm that specialty is not driving the non-linear first stage relationship initially observed. If anything, as we move closer to leveraging within-specialty variation, one might argue that the non-linear pattern become more pronounced at higher proportions of female peers of peers (i.e., above 75%). While the point estimates indicate a change in slope direction from negative to positive at the upper end of the distribution, the 95% confidence interval does not provide sufficient evidence to reject the null hypothesis of no association between the two variables.

---

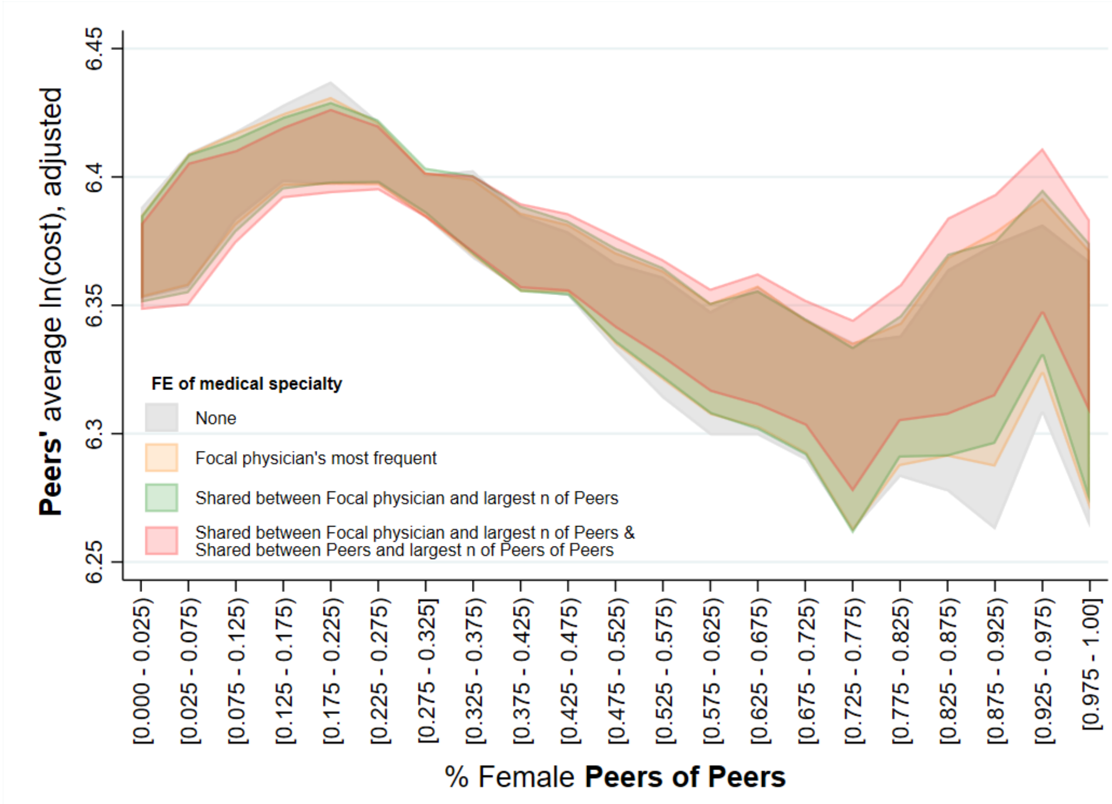
[60] For instance, if specialties with the highest hospitalization costs present a proportion of female physicians close to 20%, moving to either side of this level could indicate a shift towards lower-spending specialties.

[61] Our data shows that the number of specialties in which focal physicians are registered range from 1 to 7. As a result, the number of specialties shared with peers also vary from 1 to 7. In order for peers to be linked to peers of peers (following the criteria detailed in Section 0), they need to be registered in medical specialties excluded from the pool of specialties of the focal physician. The number of shared specialties between peers and peers of peers range from 1 to 27.

[62] We prefer that to adding the exact specialty reported at the given hospitalization as this choice could be endogenous – this is the reason why we fixed specialties at the physician level. Results are, however, not sensitive to this decision given that the specialty reported at the hospitalization level coincides with the most frequently reported specialty in 96.6% of observations.

[63] The focal physician's specialty shared with the largest number of peers differs from their most frequently reported specialty for 19% of observations.

Figure 3.4: Relationship between (adjusted) peer outcome and IV, conditional on medical specialty



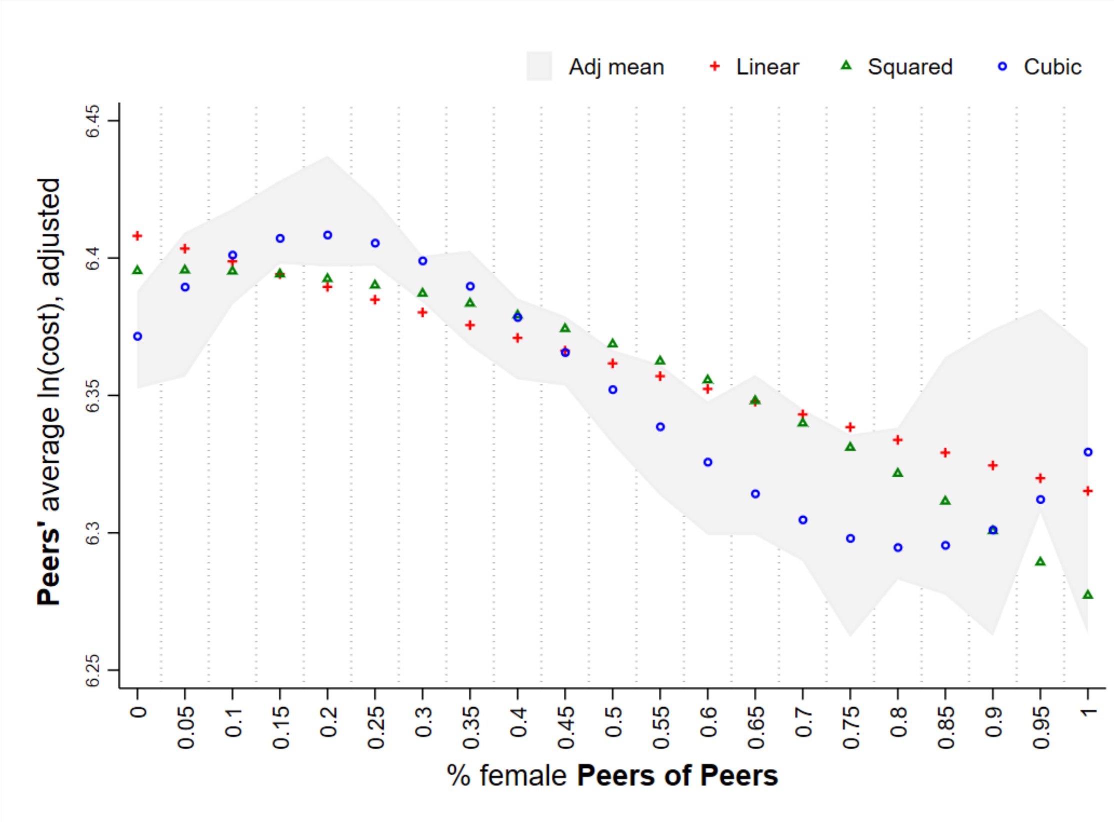
Notes: This figure illustrates the 95% confidence interval of the peer outcome adjusted mean under different specifications. Estimates presented in Figure 3.3 are replicated in light grey (i.e., baseline specification). The remaining estimates correspond to different specifications where we add different types of fixed effects for medical specialty. The estimation procedure employed to derive these adjusted means is detailed in the notes accompanying Figure 3.3 as well as in the main text.

In the appendix, Table B.1 compares raw descriptive statistics across different intervals of the proportion of females among peers of peers. We note that hospitalizations conducted by physicians whose direct peers are themselves exposed to an average share of less than 20% female peers (i.e., peers of peers) are slightly more concentrated in less socioeconomically developed regions and non-teaching hospitals, which are typically associated with lower quality. Observed differences are larger when we further break these intervals into smaller brackets, particularly when the female representation falls below 2.5%. No noticeable differences are observed in terms of hospital type (specialised vs. general) and admission process (general vs. referral only). As previously explained, our analysis controls for hospital characteristics and municipality fixed effects. Regarding physician specialty, we observe differences, although not particularly striking. This aligns with our findings indicating that accounting for variability in physician specialty does not fundamentally alter our first-stage results.

We proceed by modelling the instrument using different parametric specifications to account for the first-stage non-linearities observed earlier.[64] Below, we assess the goodness of fit of the predicted (adjusted) peer outcome variable from these first-stage specifications. Figure 3.3 plots the point estimates of each specification alongside the 95% confidence interval of the non-parametric relationship depicted in Panel A of Figure 3.3 (highlighted in light grey). As anticipated from visual examination, the cubic specification demonstrates a better fit for the first-stage relationship compared to the linear and quadratic alternatives. Figure B.2, in Appendix B, contrasts the cubic specification with linear piecewise parametric regressions featuring one knot at 0.2 and two knots at 0.25 and 0.75. The linear piecewise regression with two knots appears to fit the model more effectively than that with a single knot at 0.20. However, the cubic polynomial specification proves relatively superior, particularly in the lower end of the distribution of the female share among peers of peers. Given its superior fit to the data and the absence of a requirement for threshold selection, unlike the linear piecewise alternative, we designate the cubic parametric specification as our preferred model. Results for all specifications are presented in the subsequent tables and figures. As will be observed, once non-linearities are accounted for, the choice between non-linear parametric models does not significantly influence our estimates.

[64] For consistency, we model gender composition of peers in an analogous way as gender composition of peers of peers.

Figure 3.5: Predicted peer outcome by IV: linear, quadratic vs cubic specifications



Notes: The 95% confidence interval, displayed in light grey, is replicated from Panel A of Figure 3.3. It illustrates the adjusted mean of the endogenous variable (y-axis) for different values of the instrument (x-axis). These adjusted means stem from a non-parametric relationship between the two variables. Overlaid on this confidence interval, the figure plots markers representing the point estimates of predicted peer outcomes for different parametric specifications between the variables, as detailed in the legend. To maintain consistency, the gender composition of peers is modelled similarly to the gender composition of peers of peers (IV). The contribution from other covariates in the model is kept fixed at their average values, calculated as the mean of the product between the estimated coefficient and the observed values.

3.7 Main Results

Table 3.8 reports first-stage results for our instrumental variable, namely the share of females among peers of peers, considering different parametric specifications. Additionally, the table includes estimates for the gender composition of peers as explanatory variables. The remaining covariates specified in Equation (3.1), which are not the primary focus of the table, are omitted due to space constraints.

Peers of peers' gender is statistically significant across all specifications. However, significance alone is insufficient; instruments must be sufficiently strong to identify the effects of interest. In line with Montiel Olea and Pflueger (2013), we use 'Effective' F-statistics to assess the instrument's strength in each specification. The estimations yield F-statistics averaging around 20. As proposed by Andrews and Stock (2018), we estimate confidence sets that are robust to weak instruments. The Anderson-Rubin (AR) test is recommended for single instrumental variables, whereas the Conditional Likelihood-Ratio (CLR) test is preferred for multiple variables due to the decreased power of the AR test in this context. Despite our instrument concerning a single (average) characteristic, non-linear parametric specifications mechanically model it with multiple variables, thus falling in the category for which CLR tests are recommended. We obtain weak-iv confidence sets that are robust to clustered standard errors using Minimum Distance estimation.[65] Note that when employing weak-instrument inference, we can only interpret confidence sets, as these methods adjust confidence intervals but are unable to provide precise point estimates.

[65] These are obtained using the Stata command -weakiv- by Pflueger & Wang (2015).

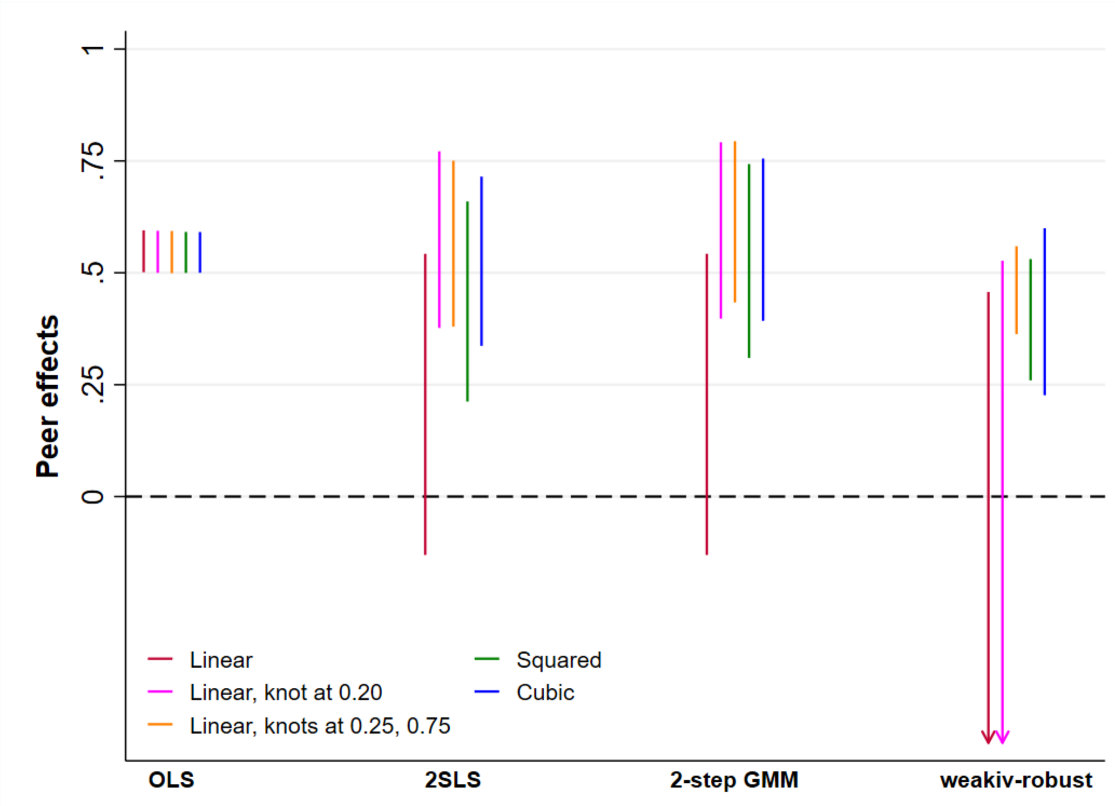
Table 3.8: First-stage results: different parametric specifications

Outcome:	Average of Ln of Peers' Hospitalization Cost				
IV:	Linear	Linear pw knot: 0.2	Linear pw knots: 0.25 & 0.75	Quadratic	Cubic
Peers of Peers' characteristics (i.e., IV)					
% female	-0.093***			0.011	0.425***
	(0.020)			(0.038)	(0.102)
% female (0.00 - 0.20)		0.202***			
		(0.062)			
% female (0.20 - 1.00)		-0.169***			
		(0.029)			
% female (0.00 - 0.25)			0.143***		
			(0.046)		
% female (0.25 - 0.75)			-0.244***		
			(0.041)		
% female (0.75 - 1.00)			0.120		
			(0.101)		
% female^2				-0.129***	-1.391***
				(0.045)	(0.294)
% female^3					0.925***
					(0.208)
Peers' characteristics					
% female	-0.344***			-0.513***	-0.272
	(0.066)			(0.093)	(0.199)
% female (0.00 - 0.20)		-0.481***			
		(0.136)			
% female (0.20 - 1.00)		-0.307***			
		(0.068)			
% female (0.00 - 0.25)			-0.424***		
			(0.113)		
% female (0.25 - 0.75)			-0.374***		
			(0.074)		
% female (0.75 - 1.00)			0.203		
			(0.167)		
% female ^ 2				0.206**	-0.522
				(0.087)	(0.471)
% female ^ 3					0.544*
					(0.324)
Diagnostic group FE	X	X	X	X	X
Municipality FE	X	X	X	X	X
Year/month FE	X	X	X	X	X
Effective F-stat	20.08	22.41	18.77	17.75	19.13
Observations	13,502,212	13,502,212	13,502,212	13,502,212	13,502,212

Notes: The distribution of the proportion of females among peers of peers, our instrument, is illustrated in Panel B of Figure 3.3 and that of the number of peers of peers contributing to this proportion can be found in Figure B.1. Estimations include all covariates and fixed effects described in Equation (3.1). Estimates for the linear specification replicate those reported in Table 3.7. For higher order polynomials, physician characteristics (not listed in the table) are modelled in a similar manner to gender. Results are similar when these variables are entered linearly in the model. The "Effective" F-statistic, following Montiel Olea and Plueger (2013), was computed using the Stata command -weakivtest-. Robust standard errors, shown in parentheses, are clustered at the municipality level. Significance levels are denoted as follows: *** p<0.01, ** p<0.05, * p<0.1.

Next, we present results for our regression of interest. Figure 3.6 displays the 95% confidence interval of (endogenous) peer effect estimates across various parametric specifications and estimation methods, including different IV estimations. Notably, while the peer outcome coefficient is not statistically different from zero under the standard linear IV specification, it becomes statistically significant when accounting for non-linearities. Peer effects uncovered by conventional IV methods - 2SLS and 2-step GMM - remain notably consistent across model specifications. Results exhibit similar patterns when weak iv-robust inference is employed, with exception of the first specifications, which become very imprecise. Overall, when considering non-linearities, our findings indicate that physicians respond to rises in peer costs by increasing their own spending by approximately *half* of the observed increase in costs by their peers.

Figure 3.6: Estimates of (endogenous) Peer Effects: baseline results



Notes: This figure presents estimates of (endogenous) peer effects for our baseline regression (i.e., ln cost as outcome, no medical specialty FE). 95% confidence intervals are outlined for different parametric specifications (linear, linear piecewise with different knots, quadratic, and cubic) and estimation methods (OLS, 2SLS, efficient 2-step GMM, and weak instrument-robust inference). Confidence sets robust to weak instrument were obtained with the command -weakiv- by Pflueger and Wang (2015), which makes use of Minimum Distance estimation and allows for clustered-robust SE. For weak-iv inference, the Anderson-Rubin (AR) test is used for the linear specification (i.e., instrumental variable represented by single variable) while the remaining specifications adopt Conditional Likelihood-Ratio (CLR) test (given that multiple variables are needed to model the non-linearities). 2SLS and (efficient) two-step GMM estimates of all the model's coefficients are presented in Table 3.9 and Table B.2.

The figure also displays OLS estimates for comparison. As anticipated, IV estimators consistently exhibit lower precision than OLS, as they exploit more constrained variation (i.e., only that induced by the first stage). Later in this section, it will become clearer why OLS estimates show minimal change when accommodating non-linearities.

Table 3.9 reports all 2SLS estimated coefficients, which closely resemble coefficients obtained through the efficient 2-step GMM method, as documented in the appendix (Table B.2). The coefficients on peer outcome, approximately 0.5, suggest that a standard deviation increase in peer average cost triggers a rise in the focal physician's own cost of around 0.20 to 0.25 standard deviations.[66]

In addition to the (endogenous) peer effects, the table also presents the direct effect of peer characteristics on physician outcomes (i.e., contextual peer effects). While peer characteristics may directly influence physician behaviour, such effects are statistically significant only for certain attributes. Specifically, estimates for gender and residency are statistically different from zero, with gender being more influential - a result consistent with our choice of instrument. We find marginal effects of opposite signs for these two characteristics. At the sample average values, 2SLS estimates indicate marginal effects of -0.29 for the share of female peers and +0.20 for the proportion of peers with completed

[66] Consider the point estimate for the cubic specification (last column of Table 3.9) of 0.526. A 1 SD increase of the ln cost of peers (0.48 as shown in Table 3.3) therefore leads to a 0.25 increase in the ln cost of the focal physician (0.48*0.526), corresponding to 0.24 of its sample SD (of 1.06, as informed in Among the focal hospitalizations in our sample, the average duration was 8 days, with an average total cost of R$1,369 (equivalent to approximately US$340 as of December 2019). This cost encompasses various expenses, including fees for medical interventions, ICU utilization, and auxiliary services such as additional diagnostic tests and consultations conducted during the hospital stay. The average age of patients is 39 years, with females constituting 60% of the total. Nearly 90% of hospitalizations conclude with patient discharge, while the remaining cases involve either patient death or transfer to other facilities. Notably, 83% of hospitalizations occur in general hospitals, with 84% taking place in facilities offering general admission processes, including both spontaneous and referred patients. Additionally, 45% of these hospitalizations occur in teaching hospitals.

Table 3.1). Linear piecewise specifications yield a similar figure of 0.26 SD.

residency degree (Table 3.9), while two-step efficient GMM estimates point to -0.25 and +0.19, respectively (Table B.2).[67] Conditional on peer behaviour and other characteristics, variations in peer composition regarding age and proportion of top university graduates do not lead to statistically or economically significant changes in physician spending patterns.

[67] The marginal effect of peer gender in the cubic specification is calculated as (-0.825) + 2*(0.943)*X + 3*(-0.331)*X^2 where X represents the proportion of female peers at which the effect is evaluated. Substituting X=0.34 (average sample proportion, as shown in Table 3.3), we find an estimated marginal effect of -0.29. Analogously, for residency, the marginal effect at the average is computed as (-0.063) + 2*(0.785)*0.55 + 3*(-0.657)*0.55^2, resulting in 0.20. These calculations are based on the 2SLS coefficient estimates presented in Table 3.9.

Table 3.9: 2SLS estimates: different parametric specifications

Outcome:	Ln of Hospitalization Cost				
IV:	Linear	Linear pw knot: 0.2	Linear pw knots: 0.25 & 0.75	Quadratic	Cubic
Peers' average outcome	0.206	0.574***	0.565***	0.436***	0.526***
	(0.171)	(0.101)	(0.095)	(0.114)	(0.096)
Peers' characteristics					
% female	-0.374***			-0.727***	-0.825***
	(0.082)			(0.117)	(0.158)
% female (0.00 - 0.20)		-0.696***			
		(0.116)			
% female (0.20 - 1.00)		-0.136***			
		(0.046)			
% female (0.00 - 0.25)			-0.626***		
			(0.095)		
% female (0.25 - 0.75)			-0.123**		
			(0.055)		
% female (0.75 - 1.00)			-0.020		
			(0.093)		
% female ^ 2				0.519***	0.943***
				(0.098)	(0.344)
% female ^ 3					-0.331
					(0.223)
average age	0.002	0.002	0.002	-0.002	-0.035
	(0.002)	(0.001)	(0.001)	(0.006)	(0.032)
average age ^ 2				0.000	0.001
				(0.000)	(0.001)
average age ^ 3					-0.000
					(0.000)
% top university degree	-0.035	-0.019	-0.019	0.063	0.080
	(0.035)	(0.024)	(0.023)	(0.050)	(0.119)
% top university degree ^ 2				-0.093*	-0.152
				(0.048)	(0.356)
% top university degree ^ 3					0.045
					(0.249)
% residency degree	0.205***	0.107***	0.105***	0.393***	-0.063
	(0.047)	(0.032)	(0.034)	(0.078)	(0.168)
% residency degree ^ 2				-0.240***	0.785**
				(0.071)	(0.379)
% residency degree ^ 3					-0.657**
					(0.260)
Focal's characteristics					
female	-0.069***	-0.069***	-0.068***	-0.068***	-0.068***
	(0.008)	(0.010)	(0.010)	(0.009)	(0.009)
age	0.001**	0.000	0.000	0.003	0.017
	(0.000)	(0.000)	(0.000)	(0.002)	(0.011)
age ^ 2				-0.000	-0.000
				(0.000)	(0.000)
age ^ 3					0.000
					(0.000)
top university degree	-0.016**	-0.017***	-0.017***	-0.016**	-0.016***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
residency degree	0.056***	0.057***	0.057***	0.055***	0.048***
	(0.021)	(0.018)	(0.018)	(0.018)	(0.016)

Outcome:	Ln of Hospitalization Cost				
IV:	Linear	Linear pw knot: 0.2	Linear pw knots: 0.25 & 0.75	Quadratic	Cubic
residency degree, ever	0.033	0.024	0.024	0.028	0.031**
	(0.020)	(0.017)	(0.017)	(0.017)	(0.015)
multiple specialty	0.074***	0.080***	0.080***	0.078***	0.078***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
employment: autonomous	0.055***	0.048***	0.049***	0.050***	0.047***
(vs staff)	(0.020)	(0.012)	(0.012)	(0.015)	(0.013)
employment: other	0.116**	0.100***	0.100***	0.106***	0.104***
(vs staff)	(0.049)	(0.034)	(0.034)	(0.039)	(0.034)
Patient					
female	-0.034***	-0.030***	-0.030***	-0.030***	-0.030***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
age	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Health facility					
teaching status	0.145**	0.075***	0.078***	0.103***	0.084***
	(0.057)	(0.029)	(0.028)	(0.038)	(0.030)
general admission protocol	0.055*	0.053**	0.053**	0.054**	0.054**
	(0.033)	(0.024)	(0.023)	(0.025)	(0.023)
type: specialised	0.028	0.028	0.028	0.029	0.028
(vs general hospital)	(0.025)	(0.020)	(0.020)	(0.020)	(0.019)
type: other	0.065	0.046	0.044	0.047	0.046
(vs general hospital)	(0.106)	(0.065)	(0.065)	(0.076)	(0.067)
Diagnostic group FE	X	X	X	X	X
Municipality FE	X	X	X	X	X
Year/month FE	X	X	X	X	X
Observations	13,502,212	13,502,212	13,502,212	13,502,212	13,502,212

Notes: Peer's average outcome is instrumented with share of females among peers of peers. For consistency, in higher order polynomial specifications (last two columns), we modelled other physician characteristics in the same way as gender. Results are very similar when we model these variables linearly. For linear piecewise regressions, we restrict the knot(s) solely to physician gender. This is because the exact knots were specifically defined for this variable (i.e., female representation), as pointed both by our first-stage non-parametric relationship investigation (as seen in Figure 3.3). Standard errors are clustered at the municipality level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Above, we examined how focal physicians respond to observed changes in female peer representation, regardless of peer behaviour, as well as to variations in peer outcomes. To compare the effects driven by peer characteristics (contextual effects) to those triggered by peer behaviour (endogenous effects or behavioural spillover), we compare the effects resulting from a marginal variation in the proportion of females among peers. For contextual effects, this is directly determined by the estimates of δ in Equation (3.1), indicating a marginal effect of -0.29, as described in the last paragraph. To assess the indirect impact through behavioural spillovers that such marginal change in the gender composition of peers would have on the outcome of the focal physician, we multiply the coefficient of peer gender ratio on peer outcome in the first-stage regression (i.e., the effect of gender composition of peers on their outcomes) by the marginal effect of peer outcome on own outcome, informed by the regression of interest ($\boldsymbol{\beta}$ in Equation (3.1)). By doing so, we find that a marginal increase in the share of female peers (at the average) would lead to a decrease in individual outcomes of -0.23.[68] Therefore, the direct and indirect effects are of similar magnitude. The overall peer effects resulting from such a shift in peers' female representation, at the sample mean, amount to (-0.29) + (-0.23) = -0.524, corresponding to an 8% decrease in focal physicians' average outcomes.

Table 3.10 illustrates the marginal effects within each interval bracket of the share of female peers, separately for direct and indirect effects along with overall effects. Notably, direct effects exhibit greater magnitude in the lower end of the distribution, whereas indirect effects become predominant when peer female representation exceeds 40%. Across most of the distribution, peer effects tend to be negative. The final column of the table displays first-stage effects resulting from marginal increases in the proportion of females among peers of peers within each level of this distribution. It's noteworthy that the positive relationship observed for low shares of female representation in the first stage is absent in the main regression of interest.[69] Despite encountering first-stage non-linearities, our main regression

[68] The marginal effect of an increase in the proportion of peer females on peer outcome is calculated as (-0.272) + 2*(-0.522)X + 3(0.544)*X^2, , based on estimates provided in Table 3.8 (first-stage regression). Substituting X with the average sample share of female peers (0.34, as shown in Table 3.3), yields -0.438. This value is then multiplied by 0.526, the estimate of β.

[69] Effects in the first- and second- stages converge, as their magnitude attenuates for higher female shares (positive first-stage effects become less positive and negative second-stage effects become less negative). Notably, when female representation equals 20%, the signs of both effects are the same. At 55%, the magnitude of the effects become identical. Interestingly, this aligns with the point where the magnitude of indirect effects exceeds that of direct effects in the second-stage coefficients.

of interest highlights peer effects that are predominantly linear with respect to the female representation among peers.

The presence of non-linearities solely in the first-stage relationship sheds light on why the estimates presented in Figure 3.6 are sensitive to the choice between linear versus non-linear specifications for IV estimation methods, whereas they remain unaffected when estimated using OLS. While the former exclusively relies on first-stage variation, where non-linearities are particularly important, the latter leverages unrestricted variation.

Table 3.10: Marginal effects of share of female peers on physician outcome

% female Peers	IV results						**% female PoP**	1st stage
	2SLS			GMM				
	overall	direct	indirect	overall	direct	indirect		overall
0	-0.97	-0.83	-0.14	-0.91	-0.76	-0.16	**0**	0.43
0.05	-0.90	-0.73	-0.17	-0.85	-0.67	-0.18	**0.05**	0.29
0.1	-0.84	-0.65	-0.19	-0.79	-0.58	-0.21	**0.1**	0.17
0.15	-0.77	-0.56	-0.21	-0.73	-0.50	-0.22	**0.15**	0.07
0.2	-0.71	-0.49	-0.22	-0.67	-0.43	-0.24	**0.2**	-0.02
0.25	-0.64	-0.42	-0.23	-0.61	-0.36	-0.25	**0.25**	-0.10
0.3	-0.58	-0.35	-0.23	-0.55	-0.30	-0.25	**0.3**	-0.16
0.35	-0.52	-0.29	-0.23	-0.49	-0.24	-0.25	**0.35**	-0.21
0.4	-0.45	-0.23	-0.23	-0.43	-0.19	-0.25	**0.4**	-0.24
0.45	-0.39	-0.18	-0.22	-0.38	-0.14	-0.24	**0.45**	-0.26
0.5	-0.33	-0.13	-0.20	-0.32	-0.10	-0.22	**0.5**	-0.27
0.55	-0.27	-0.09	-0.19	-0.27	-0.06	-0.20	**0.55**	-0.27
0.6	-0.21	-0.05	-0.16	-0.21	-0.03	-0.18	**0.6**	-0.25
0.65	-0.16	-0.02	-0.14	-0.16	-0.01	-0.15	**0.65**	-0.21
0.7	-0.10	0.01	-0.11	-0.11	0.01	-0.12	**0.7**	-0.16
0.75	-0.04	0.03	-0.07	-0.05	0.03	-0.08	**0.75**	-0.10
0.8	0.02	0.05	-0.03	0.00	0.03	-0.04	**0.8**	-0.02
0.85	0.07	0.06	0.01	0.05	0.04	0.01	**0.85**	0.07
0.9	0.13	0.07	0.06	0.10	0.04	0.06	**0.9**	0.17
0.95	0.18	0.07	0.11	0.15	0.03	0.12	**0.95**	0.29
1	0.23	0.07	0.17	0.20	0.02	0.18	**1**	0.42

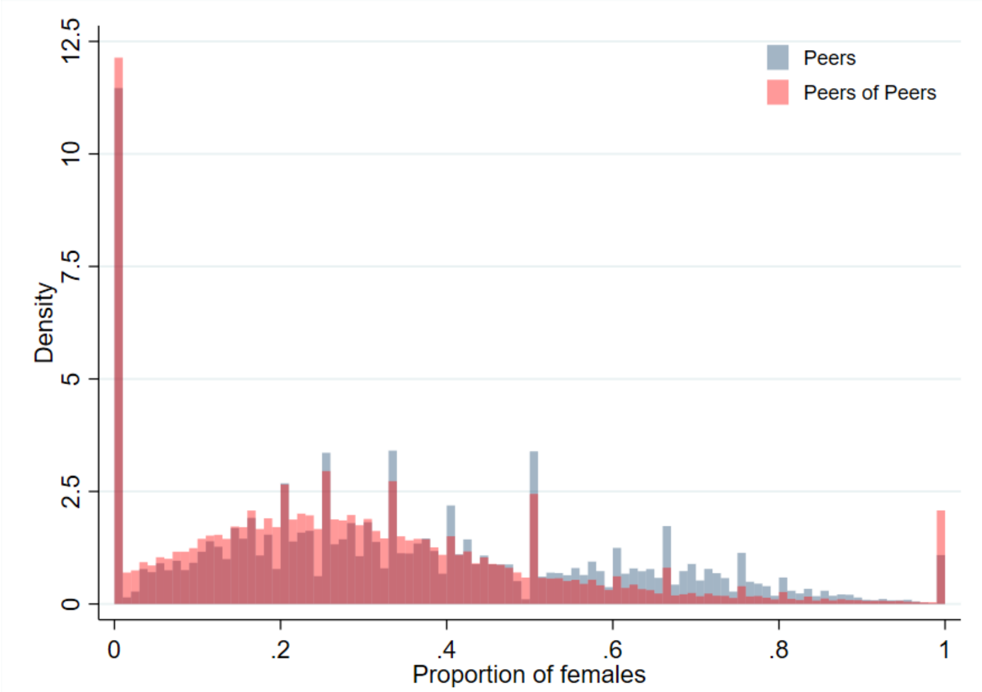
Notes: This table shows marginal effects of the proportion of female peers to which physicians are exposed on their behaviour. Outcome is measured as the natural logarithm of hospitalization cost. Marginal effects are computed based on coefficient estimates from the cubic specification reported in Table 3.9 (2SLS), Table B.2 (GMM), and Table 3.8 (first-stage). The columns entitled "direct" correspond to contextual peer effects (i.e., direct effect of peer characteristics on focal physician's outcome). To compute these, we resort to the coefficient estimates of female share of peers in the main regression results (Table 3.9 for 2SLS, Table B.2 for 2-step GMM). The columns entitled "indirect" correspond to behavioural spillovers on focal physician outcome from changes in peer outcome that is triggered by changes in gender composition as described in the first column. To compute these, we use both coefficient estimates of female share of peers in the first-stage regression (Table 3.8) as well as coefficient estimates of the peer outcome variable in the main results (Table 3.9 for 2SLS, Table B.2 for 2-step GMM).

One likely reason for the dissimilar functional forms between the first- and second-stage regressions is the utilization of distinct populations of focal physicians, peers, and peers of peers in our estimation. A number of important points should be stressed. Firstly, our estimation sample consists of focal physicians who, by design, exhibit lower network centrality compared to their direct peers. This is because we specifically select focal physicians employed solely within the given hospital (around the time of hospitalization) and who tend to specialize in fewer medical fields. The latter stems from the naturally higher likelihood of finding associated non-empty sets of peers of peers in specialties excluded from the focal physician's own set when this set is small. In contrast, peers generally have higher network centrality if they are linked with non-empty sets of peers of peers, which requires employment across multiple hospitals and registration in additional medical specialties beyond those shared with the focal physician. Secondly, we impose the condition that the single hospital where the focal physician works is government-owned. As previously mentioned, this restriction is justified by our ability to observe all activities occurring within these types of facilities. Conversely, we impose no restrictions on the ownership of other hospitals where peers of focal physicians simultaneously practice. Therefore, our set of peers of peers consists of physicians working alongside peers in any hospital providing services to SUS, including private institutions.

Figure 3.7 highlights notable distinctions in the distributions of the proportion of females among peers (in blue) and peers of peers (in red) within our estimation sample. Firstly, the distribution of female shares among peers is shifted to the right of that of peers of peers (with averages of 0.34 and 0.29 respectively, as detailed in Table 3.3). Secondly, a substantial concentration of observations around 0.2 is observed in the peers of peers' distribution, aligning precisely with the range where the first-stage non-linearity was identified. The higher density of observations for proportions of female representation between 0 and 0.2 among peers of peers facilitates the depiction of non-linear relationships within this specific range. Conversely, the distribution among peers exhibits a more uniform spread between 0 and 1, with notable peaks at certain values (e.g., 0, 1, 1/2, 1/3, 2/3, etc). These peaks are much more pronounced in this distribution because the number of peers among which the shares are computed tends to be much lower than that of peers of peers (see Panels B and C in Figure B.1).

Moreover, it's crucial to acknowledge a conceptual difference between the first- and second-stage estimates. While the latter provides estimates of peer effects for the focal physician population, the former cannot be used to recover peer effects for the population of peers. The reason for this is that our estimation method uses only a subset of the peers' own peers to construct our instrument, specifically those meeting the exclusion restriction criteria. To consistently estimate peer effects, all peers to whom the physician was exposed should be included in the right-hand side of the equation.

Figure 3.7: Distribution of female shares among peers and peers of peers



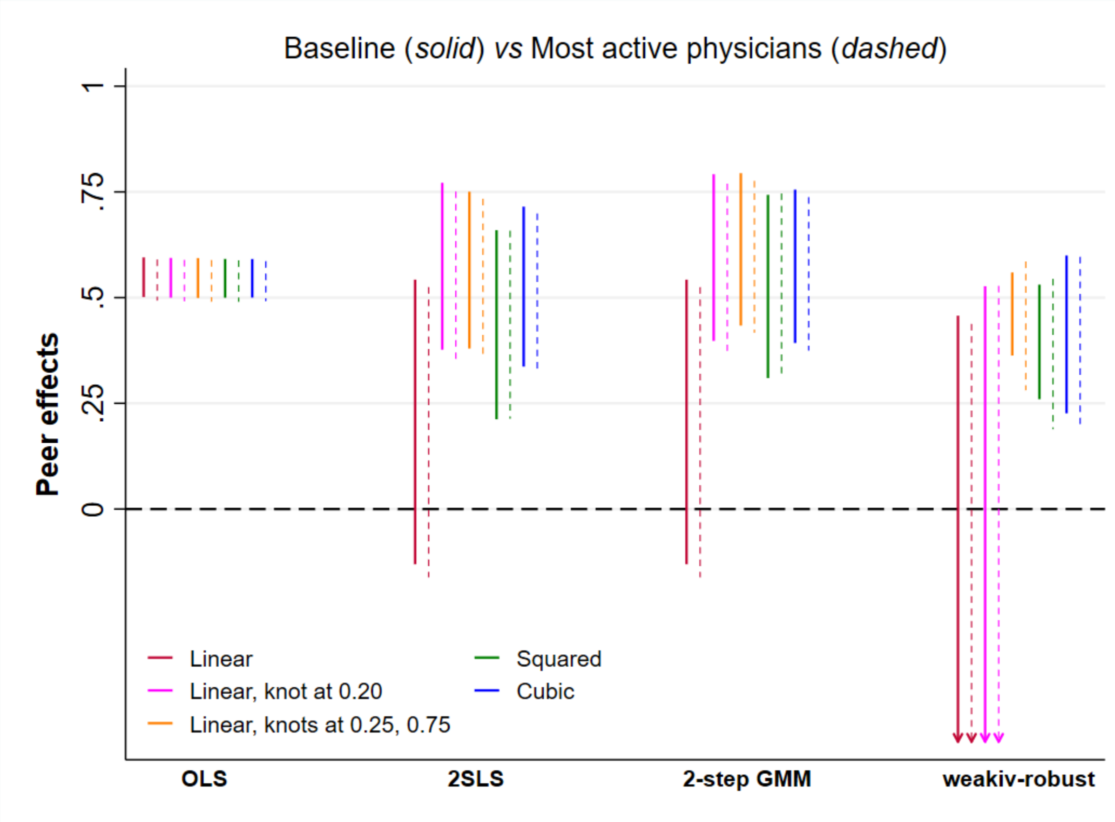
Notes: For each observation in our estimation sample, we observe share of females among Peers and share of females among Peers of Peers (i.e., instrument). The figure overlays the distribution of the latter (in red, already presented in Panel B of Figure 3.3) on top of the former (in blue). The peaks stem from the fact that, for small groups, shares tend to concentrate in specific values. For instance, for groups of single individuals, the share is either 0 or 1; for groups of 2 individuals, the possible values of shares are 0, 0.5, or 1; and so on. Because, for a given observation, the number of associated Peers of Peers tend to be larger than that of Peers (i.e., they include peers of peers for each one of these peers), the distribution is smoother.

3.8 Robustness analyses

Our findings suggest that physicians respond to observed variation in the behaviour of nearby peers by incorporating approximately half of this change to their own behaviour. This result, as demonstrated in Figure 3.6, remains robust to weak instrument inference and different parametric models that accommodate non-linearities. In this section, we conduct two supplementary analyses. First, we estimate results for the sample of high-volume physicians to investigate whether our baseline findings are influenced by occasional health providers who may exhibit outlier hospitalization outcomes. Second, we leverage more granular variation by accounting for heterogeneity across medical specialties.

In the first analysis, we restrict our final estimation sample to hospitalizations conducted by physicians identified in the unrestricted data as having a total number of hospitalizations above the 50^{th} percentile. This corresponds to 108 hospitalizations throughout the 7.5 years of our analysis period. Although caution should be taken when interpreting this sample,[70] the analysis is deemed informative as it leverages data from physicians who are observed more often in the dataset, thus mitigating concerns regarding idiosyncratic patterns and recurring mismeasurements. Notably, the results remain almost identical, as illustrated by Figure 3.8, where the dashed lines represent the restricted sample.[71]

Figure 3.8: Estimates of (endogenous) Peer Effects: most active physicians



Notes: Solid lines replicate the baseline results presented in Figure 3.6. In dashed lines, we present the corresponding 95% confidence interval for results obtained for the subsample of physicians above the 50th percentile in total number of hospitalizations between July/2012 and December/2019, which corresponds to 108 hospitalizations. All notes of Figure 3.6 apply.

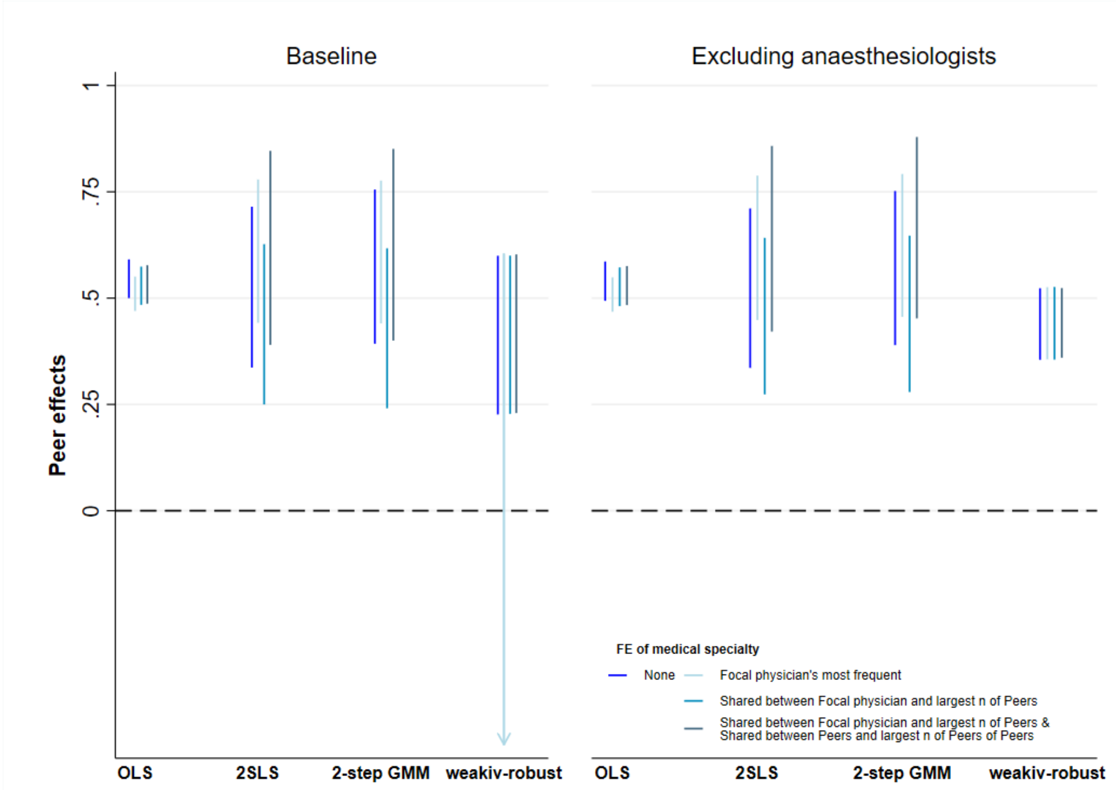
Lastly, we provide evidence that our estimates are not driven by heterogeneity in medical specialties. Figure 3.9 displays peer effect estimates while accounting for medical specialty fixed effects measured through several approaches, which were used earlier to

[70] Physicians with high hospitalization volumes may, to some extent, be more likely to perform simpler procedures, typically less time-consuming. Moreover, by limiting the estimation sample to physicians with a high total number of hospitalizations, we risk excluding those who are significantly active in their specific procedures or younger physicians who entered the sample later.

[71] The unrestricted dataset concerns that comprising the universe of all hospitalizations by the initial sample of physicians (described in Section 3.3.4). The minimal alteration to our estimation sample when we perform this restriction indicates that our baseline sample primarily consists of hospitalizations by high-volume physicians. Specifically, the 1% least active physicians in our baseline sample correspond to those in the 10th percentile in the unrestricted data, and the 10% least active are located in the 24th percentile before implementing the restrictions described in Section 3.4.5.

generate Figure 3.4. Our baseline results are illustrated in royal blue, with alternative specifications shown in lighter shades of blue. Notably, the results remain highly consistent across these estimations, particularly when examining confidence sets robust to weak instrument and when excluding focal physicians who are also registered as anaesthesiologists.[72] The 95% confidence interval estimates for this group range from 0.35 to 0.52.

Figure 3.9: Estimates of (endogenous) Peer Effects: conditional on medical specialty



Notes: In royal blue, we replicate estimates of the cubic parametric model in our baseline results presented in Figure 3.3. In other shades of blue, we show the corresponding 95% confidence interval after adding medical specialty FE to the regression model. There are three different ways to account for medical specialty. First, we consider the most frequently reported specialty by the focal physician. An alternative is to include FE of the specialty shared between the focal physician and the largest number of peers. Because physicians may hold more than one medical specialty, this is not always the same as the prior. Finally, in addition to the latter, we also add the peers' medical specialty that is most commonly shared with peers of peers. The plot on the right excludes from our estimation sample 1,420 focal physicians for whom anaesthesiology is among their set of medical specialties. This reduces the number of observations in our estimation sample from 13,502,212 to 13,206,686 observations. All notes of Figure 3.6 apply.

72 As detailed in Section 3.4.5, although we excluded hospitalizations by physicians who reported having provided anaesthesiology services during a given episode, our baseline estimation sample kept records where these physicians reported having provided services in other medical specialties. In the right-hand side plot of Figure 3.9, we excluded all physicians who ever reported having worked as anaesthesiologists, as they may present very particular care patterns.

3.9 Discussion

In this chapter, we examine the influence of peers on physicians' treatment behaviour within the hospital setting. Physicians' treatment decisions are shaped by their beliefs, preferences, and skills acquired through formal training and on-the-job experience. Working alongside colleagues with diverse treatment styles offers physicians opportunities to update these determinants through knowledge exchange and norm adherence.[73]

Physician behaviour is evaluated through the assessment of their spending patterns. Hospitalization costs are determined according to the national fee schedule. They comprise procedures conducted, tests ordered, auxiliary services provided, and the utilization of resource-intensive hospital facilities (e.g., intensive care unit) during a patient's hospital stay. While costs are a salient metric of providers' decision-making in patient care, they are not directly linked to financial incentives for physicians, who typically receive fixed salaries or are employed on a shift basis, nor for hospitals, which generally operate on fixed budgets predetermined by past activity and existing infrastructure. Moreover, costs are not borne by patients, who have free access to care at the point of use.

To study this question, we leverage rich administrative data enabling us to map the entire network of roughly 200,000 physicians within the public healthcare system. This data source is matched with physician-level registries containing detailed information on demographics and educational backgrounds. Network intransitivity allows us to find peers' links with physicians in medical fields and health facilities which are excluded from the network of the focal physician (i.e., peers of peers). Shocks to the composition of this group lead to shifts in their characteristics and behaviours, which prompt peers to adjust their behaviour in response to peer effects. These adjustments in peer behaviour subsequently influence focal physicians to adapt their own practices accordingly.

In an IV approach, we use gender composition of those peers of peers to instrument the average costs of peer hospitalizations within the past 30 days. Besides the latter being a sufficiently relevant predictor of the costs incurred by peers, two additional assumptions are necessary for the causal identification of peer effects. The exclusion restriction assumption requires that the gender ratio of peers of peers only influences the physician's own costs

[73] Treatment choices based on norm compliance may, at first, not interact with any of these determinants. However, over time, it is expected to do so given that physicians acquire experience from adopting the new treatment options.

through its effect on the recently incurred costs of peers, conditional on the physician's own attributes as well as those of the peers. Heterogeneity in diagnosis, patient characteristics, hospital observables, as well as municipality- and time-specific effects are taken into account. The exchangeability assumption requires the absence of factors influencing both the physician's hospitalization and the gender ratio of peers of peers. The validity of this second assumption could be called into question in the event of unexpected disruptions to the practice environments of both the focal physician and the selected peers of peers.

Shocks to the hospital or medical specialty of the peers of peers that affect their female representation do not threaten our results as long as they are not correlated with shocks to the hospital or specialty of the focal physician. If such correlations do exist but are absorbed by variations in the composition of the focal physician's immediate peers, then the identification assumptions are still satisfied. Our results would remain valid if, for instance, common shocks to hospital funding prompted both institutions to adjust their medical staff composition similarly. Analogously, if an innovation introduced in both medical fields triggered similar types of physicians to become more active in both fields, our findings would remain valid.

Our findings indicate that working alongside more resource-intensive peers drives physicians to increase their use of medical inputs when treating their own patients. We find that doctors incorporate approximately *half* of the observed changes in peer spending. Results remain robust to restricting our sample to high-volume physicians and exploiting within-specialty variation. Our estimation sample suggests that a one standard deviation increase in peer average cost leads to a 0.24 standard deviation increase in own cost.

It is not straightforward to compare our results to other evidence in the literature given differences in model specifications, measured outcomes, and the type of care under evaluation. We believe the closest paper to ours to be Barrenho et al. (2023), who look at innovation take-up in the UK and find that an increase of one standard deviation in peer outcome results in a 0.13 standard deviation increase in focal physician outcome. The stronger effects we estimate are consistent with two different features between the studies. Firstly, our measure of costs comprises any change in physician activity, not only changes in the choice of main medical intervention as in Barrenho et al. (2023). Secondly, our peer outcome variable consists of more recent activity to which physicians are exposed, which could potentially exert a greater impact on their decisions.[74] Our findings also relate to research suggesting that switching practice environments significantly influences physicians' care patterns. Molitor (2018) demonstrates that environmental factors account for twice as much variation as physician-fixed factors in explaining practice style, while Avdic et al. (2023) attribute half of the observed variation to social factors.

Additionally, gender emerges as an influential peer characteristic. Our results indicate that an increase in the female representation of peers directly impacts physicians' spending patterns, with a magnitude largely equivalent to its indirect effects through the variations in peer spending accompanying the group compositional change. Overall, a marginal increase in the share of female peers (at its sample average) causes physicians to decrease their hospital spending by 8%. These findings corroborate the literature across various disciplines, including economics, psychology, organization, and gender studies, which consistently highlight the significance of female representation in influencing team dynamics within medical environments (Cardador et al., 2022; Sarsons, 2017; Wallace, 2014).

There are several limitations to consider when interpreting our results. First, our estimates encompass all medical specialties, potentially resulting in estimates that could be perceived as overly generic. Second, log-transformed outcomes render interpretation more challenging. Despite offering meaningful insights in percentage terms, understanding the magnitude of the results is less straightforward. Third, our instrument has proved not to be sufficiently strong for us to rely on classical inference. As recommended by the instrumental variable literature, we employ weak-instrument inference, which provides confidence sets but not precise point estimates. Finally, due to the absence of data on health outcomes, we are unable to explore the implications of peer effects on patient welfare.

While we cannot judge whether peer effects in health spending are beneficial or not, it's crucial to emphasize the specific context of our study. The Brazilian public healthcare system is characterized by low procedure fees and a lack of financial incentives, thus rendering induced demand improbable. In environments where under-provision of care is prevalent, there is room for health improvements through increased resource utilization. This context markedly differs from settings featuring high treatment costs and organizational

[74] In Barrenho et al. (2023), the outcome variable corresponds to the annual share of hospitalizations where the innovation was adopted. Their peer outcome variable is a cumulative measure, incorporating all previous years where physicians worked in the same hospital. In contrast, our peer outcome measure consists of the overall costs incurred by peers in the preceding 30 days.

frameworks that incentivize providers to offer unnecessary treatments. In such settings, inefficiency resulting from peer-induced spending increases would be more likely than in the context of this study. Indeed, there is evidence from US and Canada that physicians with higher skill levels generally exhibit lower spending patterns (Chan et al., 2022; Doyle et al., 2010; Gowrisankaran et al., 2022). Contrary to this, our data reveals a positive association between physician quality, as indicated by attending residency training, and spending.

Policies that influence team composition have the potential to enhance quality of care by fostering social learning. Allocative policies based on skills have been studied in the education economics literature. For instance, Leuven and Rønning (2016) found that students placed in mixed-grade classrooms tend to outperform those in more homogeneous settings. Their findings highlight notably positive outcomes for less skilled students, with diminishing returns observed for those at the higher end of the skill distribution. While concerns may arise regarding the impact of such policies on individuals with high potential, these considerations are less relevant in the context of healthcare, where maximizing patient welfare is the ultimate goal.

As our estimates are based on a linear-in-mean model, which solely capture effects in terms of peers' average characteristics and outcomes, they offer limited insight into the effects of allocative policies. To gather robust evidence regarding the effects of altering team composition at the margin, future research should consider model specifications that more precisely characterize the distribution of peers' characteristics. Besides, potential detrimental effects of team disruptions should be considered before allocative policies are introduced.[75]

One additional aspect to consider is that the magnitude of peer effects may be attenuated in settings where policymakers dictate team composition given that it is reasonable to expect that doctors may be more influenced by peers with whom they spontaneously sort. Despite the potential weakening of effects, such initiatives could prove especially beneficial in areas where social learning and knowledge dissemination would yield higher societal returns - those with a low baseline average quality of physicians.

[75] Recent research suggests that team disruptions can negatively impact provider productivity and patient outcomes, while maintaining stable teams over longer periods may offer benefits (Agha et al., 2022; Bartel et al., 2014; Chen, 2021; Stecher, 2023).

4 EFFECTIVENESS AND HEALTH IMPACTS FROM RATIONING C-SECTION USE

While variations in medical treatments among providers may not necessarily be inefficient, as extensively discussed in the preceding chapters, this is less likely to hold true when considering the choice of childbirth delivery method. There is a clear understanding that unwarranted C-sections are increasingly used. The global proportion of births delivered by C-section has risen significantly from 7.6% to 21% between 1994 and 2021, exceeding the 15% level recommended by the World Health Organization (WHO). In numerous nations, the prevalence of this surgical alternative has surpassed that of vaginal deliveries (Betran et al., 2021; Betrán et al., 2016). Factors such as advancing maternal age, higher shares of women with prior C-sections, and improved procedural safety may account for part of this upward trend (Lancet, 2000). Yet, the rapid increase and deviation from recommended rates strongly suggests that a significant portion of these procedures are driven by factors unrelated to medical need.

The accelerated trend in C-section use implies a prevailing belief that these procedures pose no harm, despite evidence indicating otherwise. Recent research leveraging plausibly exogenous variations induced by non-medical incentives reveals detrimental effects of C-sections on infant health, particularly due to respiratory disorders (Card et al., 2023; Costa-Ramón et al., 2018, 2021; Jachetta, 2016). This aligns with a well-established correlation documented in the medical literature.[76]

While the prevalence of C-sections has been on the rise, there is accumulating evidence indicating significant variations in their utilization across regions within the same country and among physicians within the same hospital (Card et al., 2023; Currie & Macleod, 2017; Epstein & Nicholson, 2009). This chapter evaluates a policy that introduced fixed constraints on the relative use of C-sections across all hospitals within the public healthcare system, *Sistema Único de Saúde* (SUS), by exploiting variation in their baseline propensity to perform the procedure.

[76] See, for instance, Davidson et al. (2010), Håkansson and Källén (2003), Hansen et al. (2008), Kristensen and Henriksen (2016), Moore et al. (2012), Roduit et al. (2009), Salam et al. (2006) Thavagnanam et al. (2008), and Tollånes et al. (2008).

In late 1990s, the Brazilian federal government introduced a cap on the proportion of reimbursable births delivered by C-section in SUS hospitals.[77] Under this directive, only up to 40% of monthly C-section procedures would qualify for reimbursement (compared to 100% previously). To safeguard hospitals from potential financial strains resulting from this cap on the relative use of C-sections, the government adjusted compensation for both types of delivery: reimbursement fees for caesarean and vaginal deliveries raised by 54% and 71%, respectively. This strategy aimed to encourage medically warranted C-sections and vaginal births while ensuring the financial stability of hospitals.[78]

Results are estimated based on a differences-in-differences research design with treatment intensity, where the latter is proxied by the extent to which the introduced cap was binding on the given municipality. Measures of exposure to the introduced threshold are constructed based on the municipality's baseline proportion of C-sections that would have gone uncompensated if the threshold had been in effect during the 12 months prior to its announcement. Estimations include municipality fixed effects, capturing time-invariant unobserved factors associated with the municipality's relative use of C-sections. Results are robust to conditioning on mother's demographics, frequency of pre-natal visits, characteristics related to C-section medical indication (such as pregnancy type and birth order), and time trends. As our analysis is performed at a more aggregate level than targeted by the policy (municipality instead of hospital level), estimates allow for within-municipality migration to hospitals less constrained by the policy.

We examine the impact of the policy on the likelihood of C-sections and health outcomes, including health at birth, hospital admissions, and mortality. The analysis of policy effects on birth outcomes relies on data from birth certificates, while estimates for mortality and hospitalization outcomes use information from death records and SUS hospital episodes. As birth certificates encompass all deliveries in the country, estimates remain robust to mothers switching to private hospitals (not targeted by the policy) within the same municipality. Death certificates cover the entire universe of fetal, infant, and maternal deaths. On the other hand, since hospitalization data is confined to the public sector, potential shifts

[77] The terms "threshold" and "cap" are used interchangeably throughout this chapter to refer to restrictions with regards to the maximum proportion of C-sections reimbursed.

[78] At the time of the policy, the average rate of C-sections across SUS hospital was 37%. C-sections were reimbursed at R$190 before the policy and R$294 after the policy, while vaginal deliveries were reimbursed at R$114 before the policy and R$195 after the policy. While increases were higher for vaginal deliveries, their fee remained lower due to the lower operational cost inherent to it.

to the private sector due to the policy could pose a challenge to interpreting our results on hospitalization outcomes. In placebo tests, evidence refuting this possibility is presented.

The results reveal a notable reduction in the likelihood of C-sections following the implementation of the policy among municipalities where the threshold was deemed more constraining. The estimates indicate that a one standard deviation increase in our measure of threshold exposure leads to an average decrease of approximately 4 percentage points in the post-policy likelihood of C-section, corresponding to 10% of the baseline mean. The effects were immediate and persisted over time. Notably, heterogeneous analyses point to higher decreases among younger mothers, first births, and deliveries from single pregnancies – instances where C-sections are less frequently medically warranted.

Reassuringly, the event study analysis supports the parallel trend assumption by revealing similar trends in the baseline period between municipalities more and less exposed to the introduced threshold. Furthermore, the absence of marked post-policy increases in C-section likelihood among municipalities where the threshold was not expected to be binding serves as strong evidence against selective migration of expectant mothers.

Results on health outcomes at birth show statistically significant decreases in the likelihood of low birthweight, implying that some C-sections eliminated by the policy might have been performed earlier than necessary. The likelihood of deliveries before the 37th week of pregnancy, however, remained unaffected by the policy. Similarly, Apgar scores (a standardized assessment of health at birth) exhibited similar trends among municipalities more constrained by the implemented threshold compared to those less constrained by it.

In terms of later health outcomes, we observe a decrease in the total number of SUS hospitalizations during the first year of life among infants born after the policy onset in municipalities where the threshold was highly binding. Quantitatively, a standard deviation increase in the SUS baseline C-section rate would result in a 1.3% decline in the total number of hospitalizations during the first year of life. This decline is driven by lower numbers of admissions, particularly within the first months of life, and is especially notable for respiratory disorders such as asthma and bronchitis. Estimates suggest that a one standard deviation rise in our measure of threshold exposure would trigger a 3.5% fall in hospital admissions caused by chronic pulmonary disorders. No effects are observed for hospitalizations of this same cohort of children due to causes presumably unrelated to the event of childbirth, thus corroborating the understanding that private sector switching is unlikely to have influenced the earlier findings.

Previous studies evaluating other interventions implemented around the world have reported limited effectiveness in reducing C-section use.[79] Among the most successful policies are those associated with decreases in C-section likelihood by less than 3.5% (Barili et al., 2021; Kozhimannil et al., 2018; C. Melo & Menezes-Filho, 2023). Moreover, evidence of unintended decreases in medically justified C-sections has been presented (Berta et al., 2020). In contrast, this chapter asserts that the Brazilian policy induces a much greater reduction in C-section likelihood, specifically limiting it to medically unjustified C-sections. While the excessive incidence of unwarranted C-sections in the baseline period could play a role, we argue that the policy design was crucial for its success.[80]

This study is closely related to Pilvar and Yousefi (2021), who assessed the introduction of a comparable threshold to government-paid C-section rates in Iran as part of a national reform. In their context, however, the threshold was set at the doctor-year level, 2 percentage points *below* the baseline average C-section rates of hospitals. While our setting shares similarities, such as the widespread use of C-section in developing countries and the type of reimbursement constraint introduced, this chapter evaluates a policy that differs in important ways. Firstly, the Brazilian threshold was set *above* the average baseline rates of targeted hospitals.[81] Secondly, it was introduced at the hospital-month level in a context where physicians received fixed compensations; thus, incentives were not directly targeted at them. Thirdly, the threshold introduction was implemented alongside generous increases in unit tariffs for both vaginal and reimbursable caesarean deliveries.[82] While implementing the cap at the hospital level aligns physician payoffs more closely with those of patients, setting

---

[79] Demand-side interventions concentrated on changes in consumer prices and access to care (Chen et al., 2014; Pilvar & Yousefi, 2021) as well as information provision (Cookson & Laliotis, 2018; C. Melo & Menezes-Filho, 2023; L. Melo, 2021). Most supply-side policies were based on changes in relative compensations of childbirth procedures (Barili et al., 2021; Berta et al., 2020; Keeler & Fok, 1996; Kozhimannil et al., 2018; Lo, 2008). While some of these policies consist of changes in direct reimbursement (i.e., fee-for-service payment models), others concern the removal of price differentials resulting from changes in the structure of payment models (e.g., replacing fee-for-service with unbundled payment, where compensation varies according to patient diagnosis instead of procedure type).

[80] Although the excessive incidence of unnecessary C-sections in Brazil could have contributed to the greater declines in C-section likelihood found in this chapter, estimates should be considered as lower bounds given that they incorporate private sector births which are not affected by the policy (i.e., those that would have happened in private hospitals regardless the event of the policy).

[81] While the baseline proportion of C-sections among all births in the country was 40.9%, the share of C-sections across SUS hospitals, targeted by the reimbursement cap, was slightly lower at 37%. The introduced cap was set at 40%, three percentage points above the average rate of targeted hospitals.

[82] The Iranian policy, on the other hand, introduced the reimbursement cap along with a bonus compensation for vaginal deliveries. At a later stage, the bonus payment was replaced by raises in procedure fees. In addition to supply side incentives, the policy also removed patient costs for vaginal births.

it at a sufficiently high level, along with increased unit fees for reimbursable childbirth deliveries, helps ensure the financial stability and quality of care of providers primarily constrained by the introduced threshold. We provide evidence that, while municipalities with a less binding reimbursement cap generally experienced higher revenue increases than those with a highly binding cap, the latter group did not incur revenue losses.

The health impacts uncovered by this study concern the causal effects of the policy. By design, the two components expected to be affected by the policy are C-section use and the amount of money reimbursed. Our finding that the decline in C-section choice occurred despite lower increases in reimbursement revenue refutes the possibility that health improvements were driven by greater availability of resources, reinforcing the understanding that results were caused by shifts in delivery type from caesarean to vaginal delivery. The positive relationship we identify between being born by C-section and future respiratory disorders finds support in the literature investigating the health impacts of C-section for children (Card et al., 2023; Costa-Ramón et al., 2021; Jachetta, 2016). Additionally, our finding of rises in birthweight following reductions in C-section is supported by research documenting the use of the surgical delivery in birth timing manipulation (Jacobson et al., 2021; C. Melo & Filho, 2021).

Finally, this study also contributes to the broad literature on geographical variation in treatment patterns. While differences in treatment dynamics have been discussed as a general topic in healthcare (Phelps, 1993), large variations have been particularly documented in the context of childbirth (Card et al., 2023; Currie & Macleod, 2017; Epstein & Nicholson, 2009). This chapter evaluates a national policy uniformly implemented across areas with substantially different propensities to perform C-sections. The observed decline in C-section use following the introduction of the reimbursement cap aligns with the well-documented understanding in the literature that supply-side financial incentives, including those not directly influencing physicians' remuneration, are important determinants of childbirth procedure choice (Foo et al., 2017; Grant, 2009; Gruber et al., 1999; Gruber & Owings, 1996).

4.1 Literature review

First, we review the literature on the determinants of C-section choice. Then, we turn to studies which evaluated different policies introduced with the aim of curbing the extensive use of C-section. Below, we use the term “provider” to refer to either physicians or hospitals given that in many contexts it is not easy to disentangle their roles.

4.1.1 Determinants of C-section choice

C-section was introduced in clinical practice as a lifesaving procedure for high-risk births. Medical indications arise from risk factors such as placenta previa (placenta implanted too close to the cervical opening), placenta accreta (placenta embedded in the uterine wall), placenta abruption (early detachment of the placenta from the uterus), premature rupture of the membranes, eclampsia, labour dystocia, foetal distress, breech presentation, and previous C-section. Medically indicated C-sections can be either planned in advance or may occur as an emergency response to complications that arise during labour.

To make sense of the accelerated adoption of C-sections over the last decades, researchers have explored the role of non-medical incentives in influencing the choice of mode of delivery. Research has shown that non-medical factors are frequently related to financial incentives and convenience motivations related to the possibility of time manipulation.

Parents and physicians have shown to resort to C-sections when they extract utility from manipulating the timing of deliveries. On the parents' side, studies suggest that cultural and financial incentives matter. Lo (2003) show that the surgical delivery is more likely to be performed on auspicious days within the lunar calendar in China, where there are beliefs that choosing the right days for important life events can change a person's fate. In contexts where child tax benefits and baby bonuses are dependent on when the child is born, several academic papers show that it is common for mothers to time their birth so as to take advantage of these financial rewards (Borra et al., 2016, 2019; Dickert-Conlin & Chandra, 1999; J. S. Gans & Leigh, 2009; Schulkind & Shapiro, 2014). Women may also prefer to deliver their babies by C-section due to perceived safety from technology-intensive procedures, fear of labour pain, and anxiety about its unpredictability and consequences to the vaginal canal.[83] De Oliveira et al. (2022) report rises in C-section use after a recent Brazilian state law increased women's autonomy to choose this type of delivery.[84]

On the providers' side, C-sections are used to manipulate the time of deliveries in order to accommodate professional commitments or due to demand for leisure. Gans et al.

[83] See Nobrega (2015) for a more exhaustive list of the reasons why Brazilian women may choose to deliver surgically.

[84] The state law was passed in Sao Paulo in the year of 2019.

(2007) document decreases in the number of births during annual obstetricians and gynaecologists' conference in the United States and Australia. Although the authors don't possess information on method of delivery, the fact that they observe an increase in the number of births just before the start of conference periods suggests that C-section were scheduled before labour onset. C-section choice motivated by the possibility to manipulate the timing of births can also occur in the labour room (i.e., no scheduling). Maibom et al. (2021) show that, during busy weeks, providers are more likely to resort to C-section to deliver births that would otherwise have happened spontaneously in ways that alleviate their workload. Other studies document a peak of unplanned C-sections at times when the opportunity cost of physicians is higher, such as end of shifts and days preceding bank holidays (Brown, 1996; Lefevre, 2014). Because labour onset is expected to be uniformly distributed in time, this is interpreted as driven by physicians' demand for leisure. An alternative hypothesis is that risk-averse physicians turn to C-sections to deliver births earlier if they anticipate deterioration in the quality of hospital services at the time natural birth would otherwise occur (e.g., after shift handover). Evidence in this direction has been shown by Fabbri et al. (2016) and Jacobson et al. (2021), who document that manipulation away from times when service quality is typically lower to be concentrated among high-risk births, which are likely to be more sensitive to the quality of care provided. Finally, risk aversion from fear of litigation could also play a role. Jachetta (2016), for instance, show that areas with higher medical malpractice premium typically experience higher rates of risk-adjusted C-sections.

In the next chapter of the thesis, we show that evidence goes in similar directions in the Brazilian context. We find that C-section is used to manipulate the timing of births away from inauspicious days (inconvenient to parents), periods when medical conferences take place (inconvenient to physicians) as well as bank holidays (in principle, times when the opportunity cost of leisure increases especially for physicians, but also times when hospital resources might be scarcer, and risk higher).

Finally, an earlier strand of the literature has focused on providers' financial motivations as an important determinant of choice for C-section. Fee-for-service payment models have typically favoured remuneration for C-section over vaginal delivery. Using US data, different papers have found empirical support for theoretical predictions that increases in Medicare's fee differential leads to more extensive use of C-sections (Alexander, 2015; Grant, 2009; Gruber et al., 1999). This has been shown to happen especially in contexts where physicians experience large asymmetry of information in their favour. Johnson and

Rehavi (2016) find that physicians respond to such incentives when treating patients with no formal medical expertise, but not when delivering babies of mothers who are doctors themselves. Furthermore, negative income shocks may lead physicians to engage more actively with the treatment option for which they receive relatively higher compensations. In a seminal paper, Gruber and Owings (1996) report evidence of obstetricians' induced demand for C-sections resulting from income losses brought by declining fertility trends in the US.

In many settings it is not clear whether changes in physicians' behaviour are driven by incentives targeted at themselves or at the hospitals they work at. Using claims data on privately insured births in California, Foo et al. (Foo et al., 2017) investigate for these separately by exploiting price variation from contract renegotiations between hospitals, physician groups, and insurers. Besides becoming more prone to perform C-sections following changes in price differential directly targeted at them, physician groups who work in single hospitals also strongly respond to price differentials benefiting the hospital that employs them. This finding is consistent with the notion that hospitals transmit their incentives for choosing specific treatment options in a way that overrides decisions solely based on medical grounds.

Finally, providers also may financially benefit from C-sections because of the shorter duration and scheduling possibility. Predictability allows for better planning and optimal use of available hospital resources, while shorter procedure timing can translate in quick turnover of delivery rooms. De Elejalde and Giolito (2021) show that higher C-section rates could offer providers higher monetary returns because of the higher number of births that can be accommodated. They refer to this as demand-smoothing mechanism. The authors advocate that this mechanism played an important role in explaining higher C-section use resulting from a policy that increased access of delivering mothers to private hospitals in Chile in a context where price incentives were not present.

4.1.2 Evaluation of policies aiming at reducing C-section likelihood

As mentioned before, fee-for-service payment models have typically compensated C-sections with higher fees than vaginal deliveries. This is because this type of procedure implies higher operational costs, involves greater risks, and requires professional expertise. Aiming at reducing the rate of unnecessary C-sections, policymakers around the globe attempted to make C-sections less financially attractive by bringing the compensated fees of both types of delivery closer together.[85]

Barili et al. (Barili et al., 2021) evaluated a regional Italian policy that equalised the tariff of caesarean and vaginal deliveries in 2005, by increasing the compensation of the latter. Despite the effect disappearing after one year, the authors show a temporal reduction of 2.6% in the C-section rate with no increase in complications of the resulting vaginal deliveries. This reduction is mainly concentrated among low-risk mothers and hospitals which are both more capacity-constrained and of lower quality, where medical decisions are more likely to be affected by factors other than health needs. As a response to this policy, private for-profit hospitals in low competitive regions have also been shown to be more likely to cut medically appropriate C-sections (Berta et al., 2020). In the same year, a similar policy was implemented in Taiwan. Lo (2008) reported no significant effect from this policy on C-section use.[86] Other papers evaluated fee equalization policies obtained by simultaneously lowering the price for C-sections and increasing the price for vaginal deliveries. Keeler and Fok (1996) find no meaningful effect from an insurance reform in California in the 1990s. Kozhimannil et al. (2018) report a reduction of 3.2% in C-sections from a similar policy introduced in the Minnesota's Medicaid program in the year of 2009.

Pilvar and Yousefi (2021) analysed the effect of a Iranian national reform that changed, in 2014, both the supply and demand incentives in favour of vaginal delivery. The policy introduced bonus payments for vaginal deliveries and a cap to the maximum annual rate of C-sections that doctors in the public sector would be compensated for. After a few months, the former was replaced with country-wide increases in the relative payment for this type of delivery. On the demand side, vaginal delivery became free for uninsured patients delivering in public hospitals. Results, which were mainly driven by the supply side, point to a decrease of 5 percentage points in the C-section rate relative to the baseline rate of 55%. The authors find that doctors for whom the quota was binding responded more significantly to the programme and, in the medium run, those with high C-section rates became more

---

[85] It is worth stressing that equalising the fees of the two childbirth procedures would translate into lower profit margin for C-sections given that this type of delivery is more costly (assuming that the agent receiving the fee-for-service compensations is the one incurring the overall costs). When providers are able to accommodate larger number of births, C-sections may still be more financially attractive given its shorter duration and possibility of time manipulation in ways that optimize resource use (i.e., physician time, physical installations, etc), as has been shown by De Elejalde and Giolito (2021).

[86] In the following year, the government added a co-payment for patients requesting C-sections despite no medical indication. Chen et al. (2014) evaluated the impact of the combined policy and found that, while the overall trend of C-section utilization did not stop, the incidence of elective C-sections was reduced.

likely to shift out of public hospitals. In terms of health outcomes, while there is evidence of increases in gestation length and birthweight, no effects were found for Apgar score, hospitalisation, and mortality.

The reform we evaluate, which was implemented in the late 1990s and targeted supply-side incentives, has not yet been assessed. Two papers evaluated another policy which was later introduced in the country. In 2016, the Brazilian Federal Council of Medicine prohibited elective C-sections to happen before 39 weeks of gestation. It also required that elective C-sections after the 39th week only be performed in mothers who expressed consent to written communication on the risks and benefits of each mode of delivery. Melo (2021) evaluated the policy using the sample of first-order births with no observed comorbidities. Based on the assumption that women's choice of procedure is only possible in the private sector, the author compared births in private hospitals (treated) with those in public hospitals (untreated), before vs after the policy implementation. He finds that, although the policy does not significantly affect the rate of C-section for low-risk first-order births, they were postponed from the 37-38th week to after the beginning of the 39th week. The same policy was evaluated by C. Melo and Menezes-Filho (2023) who used the unrestricted sample of births and a different empirical strategy. Their contrafactual group consisted of births with breech or transverse presentation of babies, which are typically interpreted as justifying a medical indication for C-section (i.e., not targeted by the policy). They find a decrease of 1.6 percentage point in the rate of C-section (roughly a 3% reduction from the baseline level of 55%), mainly driven by higher-order births in the public sector.[87] Although a slight increase in first-minute Apgar score is reported, such increase disappears by the fifth minute after birth.

Finally, other policies solely targeting cultural aspects behind the upward C-section trends have been shown not to be very effective. For instance, Cookson and Laliotis (2018) found small impacts from a government initiative aimed at promoting normal birth in England.

[87] The authors draw attention to the relevance of how the demand-side component is framed in such policy. While they show a decrease in the likelihood of C-section when mothers are allowed to make an informed choice regarding the choice of delivery method, De Oliveira et al. (2022) find that C-section increases when mothers are given autonomy to opt for a C-section without receiving information about the risks involved in their decisions.

4.2 Institutional background

4.2.1 SUS: funding and access to hospital services

Inspired by the National Health Services in the United Kingdom, the Brazilian public health system (*Sistema Único de Saúde*, SUS) was created in the year of 1990. SUS offers free healthcare at the point of use to everyone in the country and currently covers approximately 75% of the Brazilian population.[88] Management and provision of healthcare is decentralized with responsibilities passed to states and municipalities.

The federal government reimburses hospitalizations in SUS following a fee-for-service payment model. The funding process is described as follows. First, hospitals inform the local authority about all the medical procedures performed in a given month through the submission of standardized inpatient electronic forms. Local authorities, then, consolidate the data sent by hospitals in their catchment area and submit it to the federal government.[89] Finally, the federal government reimburses the local authorities for the services performed in their region according to the national fee schedule.[90] It is worth noting that, although funding between the federal and local governments are based on a fee-for-service payment model, financial arrangements between the latter and hospitals (e.g. amount and frequency of payment) are defined at the local level. Local health authorities are ultimately responsible for the financial health of its providers. If, for instance, the federal transfers are not sufficient to cover the hospital bill, they are usually required to complement the funding with their own local budgets. There has been, so far, a widespread understanding that the national schedule is outdated in regard to the fees of many procedures.[91]

While local health authorities receive funding from the federal government according to past activity, physicians in the public sector are compensated regardless of the number and type of procedures performed. Most doctors working in SUS are civil servants who receive a fixed monthly salary based on a contracted number of hours per week. Hospitals

---

[88] As of May 2023, 26% of Brazilians owned some sort of private health insurance according to the Regulatory Agency for Private Health Insurance and Plans (ANS). This figure was 17.8% back in December 2000, the first month with available information. Source: http://www.ans.gov.br/anstabnet/cgi-bin/dh?dados/tabnet_tx.def

[89] Local health authorities are typically represented by the health department of the municipal government.

[90] The total amount reimbursed must be inferior to the financial ceiling specific to the local health authority. The latter is negotiated between the federal government and the local health authority every year based on the region's demographics, past provision of hospital care, and potential plans for expansion.

[91] See, for instance, Machado et al. (2022).

providing care to SUS are also allowed to independently hire physicians, who are usually remunerated by 12- or 24-hour medical shifts.

Although access to SUS is universal, there are restrictions as to where citizens can seek care based on their municipality of residence. Local governments are responsible for the provision of medical services to their own residents. If a given service is not available in a municipality, the respective local authority is expected to have a formal agreement in place with another local authority (usually a neighboring municipality) which entitles them to direct their residents to facilities located in the outsourced municipality.

4.2.2 Policy intervention

As part of an effort to reduce the incidence of C-sections in the public sector, the federal government issued regulations imposing a fixed cap to the monthly rate of C-sections that providers could claim reimbursement for. The threshold was set at the *hospital level* and was *unique* across all facilities offering childbirth services within the public health system. The threshold was implemented into the system in a way that C-sections would get rejected once the proportion of C-sections relative to the total number of deliveries performed in the particular hospital-month surpassed the given threshold.[92]

The introduction of a 40% cap to the monthly rate of C-sections was announced on May 29, 1998, and came into force a few days later, on June 1, 1998.[93] In January 1999, the cap was updated to 37%. It was further reduced to 35% in January 2000. The fixed threshold was kept in place for all hospitals in the public system until May 2000. In June 2000, the Ministry of Health attempted to decentralise efforts to reduce C-section use to local governments by giving states the option to sign a cooperation agreement in which they would take responsibility to monitor the rate of C-sections in the region. The fixed threshold was lifted for all hospitals located in states that consented to such agreement.

Concomitant with the cap introduction to the proportion of C-section eligible for reimbursement, the national reform included other policy changes. The government

[92] Hospitals have to input inpatient electronic forms into the system every month to claim reimbursement to the federal government. After each individual form is entered into the system by a given hospital in a given month, the cumulative proportion of C-sections is automatically computed. The system only allows for new C-section claims to be entered as long as the cumulative rate for the given hospital-month is below the threshold.

[93] More specifically, the threshold was introduced into the SUS operational system in the month of July of 1998, when the system was populated with reimbursement claims of hospital episodes with discharge dates ocurring in the previous calendar month. The policy is, therefore, expected to have started its effects on hospital stays ending on June 1, 1998, a few days after its announcement.

readjusted the reimbursed fees for both normal and caesarean deliveries by more than 50%. Given higher raises to the compensation of vaginal deliveries, the relative fee reimbursed for C-sections declined from 1.66 to 1.51.[94] The federal government also introduced new reimbursable childbirth procedure codes for billing. Childbirth procedure codes for high-risk patients were added to the national fee schedule, with tariffs equivalent to twice those of the respective delivery methods. A third procedure code was created for low-risk normal deliveries performed by obstetric nurses with an associated fee corresponding to 85% of that of the vaginal delivery regular procedure code. Together with these fee adjustments, the government also authorized reimbursement of anaesthetic for vaginal deliveries (before, these were restricted to C-sections).

This chapter evaluates the effect of the national reform which imposed a reimbursement cap to the relative use of C-section among hospitals in the public system simultaneously to changes to reimbursement fees at the procedure level.

4.3 Data

We can identify deliveries in two different data sources: the national birth certificates (*Sistema de Informações sobre Nascidos Vivos*, SINASC) and SUS hospital claims (*Sistema de Informações Hospitalares*, SIH/SUS). The former dataset contains the universe of all registered live births in Brazil, whereas the latter is restricted to deliveries in SUS. More specifically, SUS hospital claims consist of administrative data on all procedures performed in SUS which were reimbursed for. Therefore, while it informs exact amounts reimbursed for childbirth deliveries by the federal government to municipalities, it does not include the universe of all deliveries taking place in SUS given that performed C-sections in excess of the threshold were not compensated for. Besides incompleteness, claims submitted after the threshold introduction could have been manipulated by providers. In reponse to the threshold implementation, derivations from the commonly known practice of upcoding could have occurred in the direction aligned with the existing incentives. As smaller compensation is better than no compensation, providers could have used the procedure code of normal delivery to bill for unreimbursable C-sections. Another possibility is that providers could have included ficticious vaginal deliveries when submitting the monthly batch of reimbursement claims to dilute the share of C-sections and receive reimbursement for all

[94] More specifically, the reimbursed fee for C-section increased from R$ 190 to R$ 294 (an increase of R$ 104, or 54%), and that of vaginal delivery rose from R$ 114 to R$ 195 (increase of R$ 81, or 71%).

performed deliveries (in addition to the ficticious ones), irrespective of the threshold. For these reasons, we will not use information that involves procedure codes from this dataset for the period after the policy implementation.

To evaluate the effectiveness of the policy in reducing the likelihood of C-section, we will resort to the SINASC dataset, whose reliability is believed to have been unaffected by the policy as its sole goal is to provide information on the country's vital statistics. Using birth certificates as our primary data source protects us against potential changes in reporting practices triggered by the policy. During our analysis period, this dataset does not inform the identifier of the hospital where births took place nor whether they were funded by SUS or the private sector.[95] Therefore, our analysis must be performed at the municipality level. The upside of using information on all births at the municipality level to evaluate the effectiveness of the policy is that it allows for spillover across hospitals located in the same municipality. This will be further explained in the next section.

Birth certificates (SINASC) include information on the exact date of birth, municipality where it took place, mother's municipality of residence, age at delivery, education level, pre-natal utilization, number of previous deliveries, indicator of multiple pregnancy, and gestational length. This data source also informs health outcomes, measured by exact weight at birth and Apgar scores at the first and fifth minutes after birth.[96]

Data on SUS hospital records (SIH/SUS) comprise all (paid) procedures, as mentioned above. It has patient-level information on demographics, such as gender, date of birth, and municipality of residence, as well as patient diagnosis, performed procedure, and dates of admission and discharge. It allows, therefore, for investigation of admissions of children born around the time of the policy as well as hospitalizations of mothers due to postpartum complication during this period. This chapter will evaluate policy impacts on these hospitalization outcomes.[97] Finally, we will also resort to this dataset to extract a

---

[95] Although, on the one hand, this is a limitation in terms of data availability, on the other hand, it is reassuring that hospitals targeted by the policy could not be monitored through this dataset.

[96] Apgar scores refer to summary measures from standardised tests performed at the 1st and 5th minutes after birth which assess the newborn's overall health based on five criteria: activity (muscle tones), pulse (heart rate), grimace (reflex irritability), appearance (skin color), and respiration (breathing rate and effort). The 1st minute score determines how well the baby tolerated the birthing process while the 5th minute score informs how well the baby is doing outside the mother's uterus.

[97] While the introduction of the C-section reimbursement cap directly affects the reliability of claims data on childbirth procedures, reporting quality of other medical procedures should have remained unaffected.

measure of exposure to the policy at the municipality level. The next section describes how this measure is calculated.

Finally, death counts are extracted from the mortality registration records (*Sistema de Informação sobre Mortalidade*, SIM). The Brazilian Ministry of Health provides distinct modules within SIM for fetal mortality, infant mortality, and maternal mortality. The first two modules include precise birth dates, with fetal mortality encompassing stillbirths and infant mortality representing deaths of children within their first year of life. Furthermore, the module dedicated to maternal mortality includes a variable indicating whether the death occurred during pregnancy or after birth. For our analysis, we focus solely on deaths that occurred after birth.

4.4 Methods

4.4.1 Empirical strategy

The aim of this chapter is to assess the effectiveness of the policy in reducing the incidence of C-sections as well as its subsequent consequences to health outcomes of infants and mothers in the first year after birth. While the evaluated reform consists of a combination of policies, this study exploits variation from its "threshold" dimension.

The research strategy employed in this study relies on a differences-in-differences model that incorporates treatment intensity as a measure of exposure to the implemented threshold (hereinafter referred to as policy exposure). This methodology is deemed more appropriate than an interrupted time series design, considering that we have information on which municipalities were expected to be affected or unaffected by the introduction of the threshold.

In our context, a uniform threshold on C-section relative use was imposed across hospitals operating within SUS, despite the wide variation in baseline propensities of hospitals to perform this procedure. Our control group will, therefore, consist of municipalities where the introduced cap was not binding in any SUS hospital, as their baseline propensity to perform cesarean procedures was lower than the level at which the threshold was set. While the treatment group comprises the remaining municipalities, our strategy leverages variation on the different levels in which the cap was deemed binding across municipalities in this group.

Our measure of policy exposure is based on the proportion of C-sections performed in the 12 months preceding the implementation of the threshold that would not have been reimbursed to the local authority (i.e., municipality) had the threshold been in place at that time. The formula presented below describes how this measure is constructed.

$$EXP_j = \frac{\sum_{H(j)} max(C_h - int(0.4T_h), 0)}{\sum_{H(j)} C_h} \qquad \textbf{(4.1)}$$

where *H(j)* refers to the set of hospitals in municipality *j* that offer childbirth delivery services within the public healthcare system, while *h* indexes each individual hospital in this set. C_h represents the number of C-sections performed in hospital *h* during the 12-month baseline period, while T_h corresponds to the total number of births delivered in hospital *h* during the same time frame. The maximum number of C-sections eligible for reimbursement is determined by $int(0.4T_h)$, which gives us (the largest integer of) the number of births corresponding to 40% of the total number of births delivered in hospital *h*. If the number of performed C-sections is lower than this maximum quantity, the numerator of the formula is zero. The denominator of the formula represents the total number of C-section births carried out in all hospitals serving the public system in municipality *j* over the 12-month baseline period. By calculating the proportion of baseline C-sections in municipality *j* that would not have been compensated due to the reimbursement cap, EXP_j measures the extent to which the implemented threshold was expected to be binding for municipality *j* had it been in place.

The decision to employ a 12-month time horizon in constructing our exposure measure to the policy is based on two considerations. Firstly, using recent baseline data enhances the reliability and representativeness of our exposure measure, providing a more accurate assessment of how binding municipalities perceived the cap at the time of its introduction. This baseline measure will be used to estimate post-policy changes caused by the cap introduction. Secondly, by adopting a 12-month annual cycle, the measure avoids contamination from seasonal effects or related idiosyncrasies. Consequently, our analysis will cover a symmetric period around the policy announcement, encompassing data from 12 months before to 12 months after. Longer periods will be examined in the event study analysis presented later in the chapter.

Using the universe of all births occurring between the 12-month period prior to the policy announcement and the 12 subsequent months, the following model is estimated:

$$CS_{ijym} = \alpha + \boldsymbol{\beta} EXP_j * Pos_{ym} + \sigma Pos_{ym} + \mu_j + \varepsilon_{ijym} \qquad \textbf{(4.2)}$$

where CS_{ijym} is a dummy that takes value 1 if delivery *i* in municipality *j* in year *y* and month *m* was by C-section, and 0 if it was a vaginal delivery. The post-policy dummy, Pos_{ym}, takes value *1* for the 12 months after the policy was announced, and 0 otherwise.[98] Municipality fixed effect, μ_j, accounts for time-invariant unobserved factors that affect the municipality's relative frequency of C-sections.[99] In extended versions of the model, we replace Pos_{ym} with indicators of year/month of birth. These monthly indicators absorb seasonal effects at the time of birth that are common to all municipalities. We also add observed characteristics, at the birth level, such as mother's age and education, frequency of pre-natal visits, type of pregnancy, and indicator of higher-order birth. The error term ε_{ijym} captures all time-variant unmeasured factors that help explain the mode of childbirth delivery. Standard errors are clustered at the municipality level.

The coefficient of the interaction term, β, is our parameter of interest. If the policy was effective, we should find $\beta < 0$. This would mean that the impact of the government reform in reducing the C-section likelihood would have been higher for municipalities where the introduced threshold represented a more binding constraint.

By evaluating the impact of the policy on the universe of all births in the given municipality, our research strategy allows for potential spillover effects across hospitals that are located in the same municipality. In response to the policy, women who have preferences for C-sections may be prompted to choose private sector facilities, where this option is more readily available. Simultaneously, women delivering in the public sector may be more inclined to select hospitals where the threshold policy was less binding. In addition, local authorities have a vested interest in allocating mothers to SUS hospitals in a way to maximize the number of compensated C-sections in their catchment area (typically delimited by the municipality). This could be accomplished by strategically assigning pregnant women, who are more likely to deliver by C-section, to hospitals that did not face binding restrictions imposed by the policy. By examining the effects of the policy at the municipality level, our findings remain robust to any movements across hospitals within the municipality.

[98] The policy was announced on May 29, 1998. Therefore, the 12-month pre-policy period corresponds to the time interval between May 29, 1997, and May 28, 1998. The analogous post-policy period elapses from May 29, 1998, until May 28, 1999.

[99] EXP_j and Pos_{ym} do not show up in the right-hand side of the regression as separate independent variables because they are absorbed, respectively, by municipality and year/month fixed effects.

Given the inclusion of municipality fixed effects in our specification, the results remain robust to factors that do not vary over time, including systematic patient characteristics and other determinants of C-section likelihood. Identification relies on the assumption that municipalities more constrained by the introduced threshold did not experience shocks that would affect the likelihood of patients delivering via C-section. The causal interpretation of our results could be compromised if, for example, mothers migrated to municipalities less constrained by the threshold to secure a C-section. We will further elaborate on this point in Section 4.7.

To assess the policy effect on health outcomes, the dependent variable in Equation (4.2) is replaced with measures of health at birth. The following outcomes are evaluated: birthweight, likelihood of low birthweight (i.e., below 2.5kg), likelihood of gestational age below 37 weeks (i.e., premature birth), magnitude of Apgar scores (which ranges from 1 to 10) at first and fifth minutes, and likelihood of low Apgar score (i.e., below 7 at fifth minute).[100] Health improvements would be primarily informed by negative estimates of β for outcomes which are highly suggestive of poor health at birth such as premature birth, low birthweight, and low Apgar score.

Given that health outcomes at birth are restricted to live births, it is crucial to assess the impact of the policy on deliveries resulting in fetal mortality. Following a similar approach as before, we compare differences in the number of deliveries resulting in stillbirths during the 12-month period before and after the policy announcement among municipalities highly exposed to the SUS threshold with analogous differences faced by municipalities where policy exposure was less important. In addition, we investigate the effects of the policy on infant death within 365 days after live birth, as well as maternal death. It is assumed that these deaths occurred within the same municipality as the childbirth delivery.

Lastly, we evaluate the impact of the policy on hospitalization counts within the public health system. This analysis is restricted to the public sector since we only have data on hospital activity for providers attached to SUS.[101] Specifically, we examine the number of

---

[100] Apgar scores lower than 7 at the 5th minute serve as a common indicator that the neonate requires further medical attention. For instance, the Neonatal Resuscitation Program guidelines use this as an indication that subsequent Apgar test assessments should be undertaken up to 20 minutes.

[101] While our sample includes all hospitalizations in the public sector, we cannot observe whether infants were born in public or private hospitals. The underlying assumption is that mothers did not become more likely to admit their children to private facilities as a response to the policy. We evaluate the plausibility of this assumption in Section 4.7.2.

hospitalizations during the first year of life among cohorts of children born in the 12-month period before and after the introduction of the threshold. Results rely on the assumption that infants are hospitalized in the same municipality as they are born in (or, more crucially, that the policy did not induce any changes on the likelihood of infants being hospitalized in their municipality of birth).[102] Additionally, we investigate the policy effects on hospitalizations of mothers due to post-partum complications.

To investigate the impact of the policy on mortality and hospitalization counts, the regression model below is estimated. The data is aggregated at the municipality-time level, where time (indexed by *t*) represents either the 12-month period preceding or following the policy announcement. A negative estimate of β would indicate that the policy triggered a fall in deaths or hospitalizations among municipalities where the introduced threshold was expectedly more binding.

$$Y_{jt} = \alpha + \boldsymbol{\beta} EXP_j * Pos_t + \sigma Pos_t + \mu_j + \varepsilon_{jt} \qquad \textbf{(4.3)}$$

Findings that the policy caused C-section likelihood to decrease and health outcomes to improve would suggest that C-sections with intrinsic negative health returns were being systematically performed prior to the reform. Evidence of health deterioration following the policy reform would imply, on the other hand, that the policy was detrimental to population health either because it resulted in cutbacks of medically justified C-sections or because the policy led to reductions in service quality during childbirth delivery (i.e., cuts to financial resources).

4.4.2 Final estimation sample

With data from SIH/SUS for the 12-month period prior to the policy announcement (May 29, 1997 - May 28, 1998), we construct our measure of policy exposure (EXP_j) following the formula described in Equation (4.1).

Policy effects on C-section likelihood and health outcomes at birth will be estimated using individual-level data. First, we select births in the SINASC database which took place

[102] This is not an unreasonable assumption given that hospitals offering treatment to infants typically offer childbirth services and mothers are likely to return to the facility where they have already received related care. Most importantly, SUS imposes restrictions as to where patients can seek care according to their municipality of residence (see Section 4.2.1). Parent relocation to other municipalities is expected to be uncommon given the short period of this analysis (up to one year after birth). As both sources of data (birth certificates and SUS hospital claims) have information on municipality of residence, we are able to empirically investigate for this (see analysis performed in Section 4.7.2).

during the 12-month period before the policy announcement (May 29, 1997 - May 28, 1998) and during the 12-month period following it (May 29, 1998 – May 28, 1999). Births with no information on mode of delivery or municipality of birth are dropped. Next, we add to the birth-level data our constructed measure of the municipality's exposure to the policy. Our final sample considers births which took place in municipalities with at least one registered birth in both the 12-month period before and after the policy as well as one C-section performed in SUS during the 12-month baseline period. We end up with 5,976,029 births (representing 95% of all births during this period) taking place in a total of 2,760 municipalities.

To estimate the effects of the policy on mortality and hospitalization outcomes, the data is aggregated at the municipality-time level, where the time dimension corresponds to birth within the 12-month period before and after the policy announcement. Using SIM data, information is gathered on the number of stillbirths, deaths of children within their first year of life, and deaths of mothers following childbirth. Additionally, counts of hospitalizations of children during their first year of life and hospitalizations due to postpartum maternal complications are extracted from SIH/SUS data. Finally, information on the municipality-level policy exposure measure is merged into the dataset.[103] The collapsed data contains information on the same sample of 2,760 municipalities.

4.4.3 Descriptive statistics

Our policy exposure measure, EXP_j, indicates that the implemented policy was expected to be binding in 1,265 municipalities (46%), while it was deemed non-binding in the remaining 1,495 municipalities (54%).[104] Figure 3.1, Panel A, plots the distribution of the baseline C-section rate in the public health system, at the level of the municipality, for these two groups. Calculating C-section rates at the municipal level is equivalent to computing the average of the C-section rates across hospitals within that municipality, considering the number of births delivered in each hospital as a weighting factor.

Non-binding municipalities consistently exhibit shares of C-sections within SUS below 40% as all SUS hospitals in these municipalities have rates below this threshold.

[103] This last merge assumes that infants are born in the same municipality as they are admitted to the hospital (see footnote 102 for credibility of this assumption). Analogously, it assumes that deaths occur in the same municipality as childbirth delivery.

[104] The threshold policy was expected to be binding in municipalities where EXP_j was positive, and non-binding in municipalities where EXP_j is equal to zero.

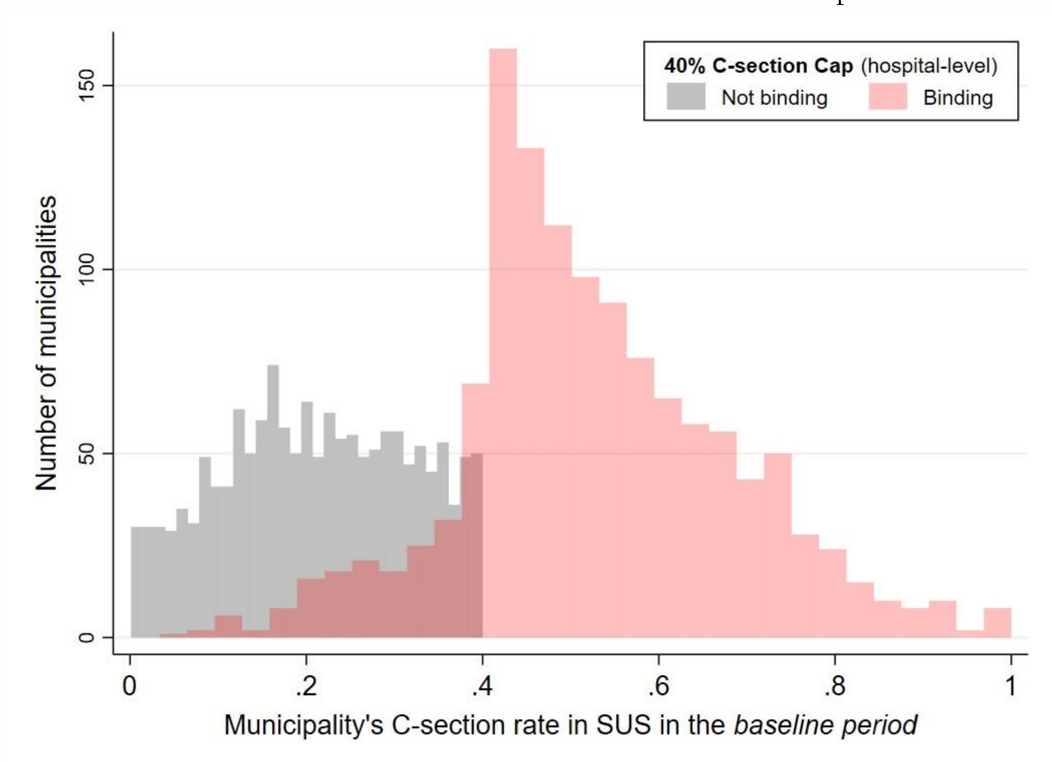
Conversely, binding municipalities are those where at least one hospital in the catchment area presents a baseline C-section rate in SUS surpassing the 40% threshold. Some municipalities within this group have shares of C-sections lower than 40%, as displayed in Panel A of Figure 3.1. This can occur whenever there are other hospitals within these municipalities that present rates below the 40% threshold.[105]

As explained in Section 4.4.1, EXP_j represents the proportion of C-sections conducted during the 12-month baseline period that would not have been reimbursed if the threshold policy was in effect. The measure, whose average and standard deviation are 0.11 and 0.16, has value zero for municipalities where all hospitals reported a baseline C-section rate below the 40% threshold. For the group of municipalities where the threshold was deemed binding (i.e., those assigned a positive measure), Panel B of Figure 3.1 plots the policy exposure measure (in the y-axis) against the municipality's baseline C-section rate in SUS (in the x-axis). The plot further delineates the quartiles of the distribution of EXP_j. The median indicates that, in half of the municipalities where the threshold was binding, 22% of the C-sections performed in SUS would have remained uncompensated had the reimbursement cap been active during the 12-month baseline period.

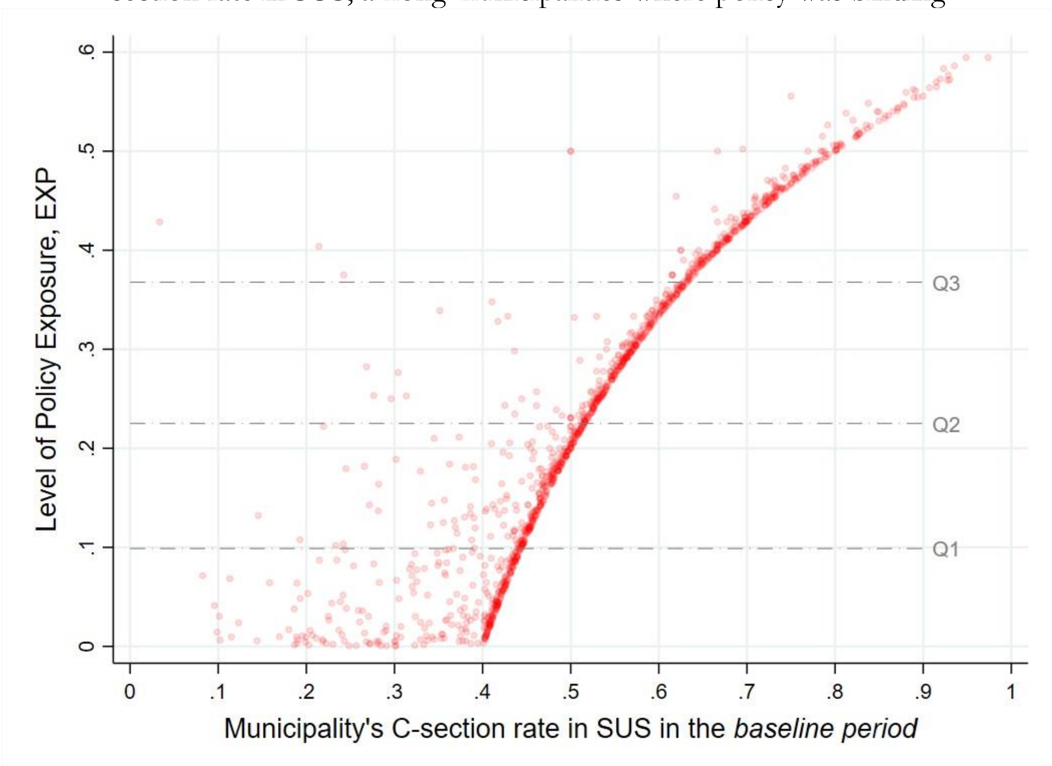
105 Municipalities where the policy was expectedly binding tend to have more hospitals providing childbirth delivery services within SUS compared to those where the policy was deemed non-binding. Specifically, 37% of municipalities in the binding group have more than one hospital offering childbirth deliveries in SUS, while this ratio is 22% for municipalities in the non-binding group.

Figure 4.1: Municipalities' level of exposure to the threshold policy

Panel A: Baseline C-section rate in SUS across all municipalities



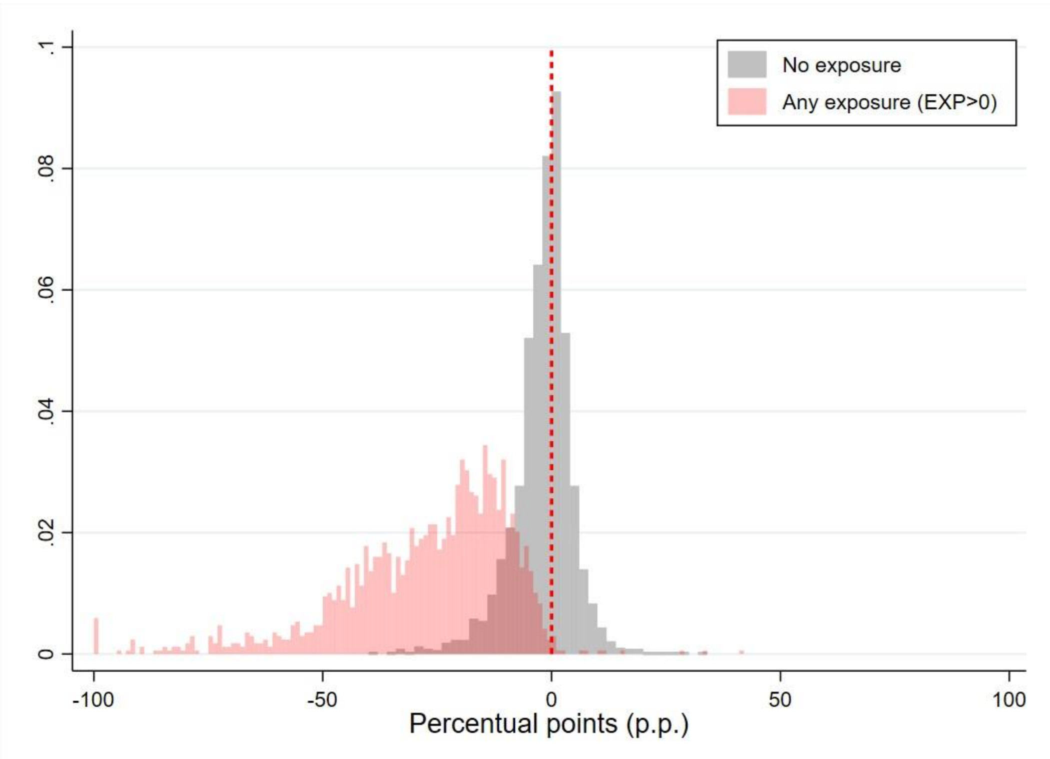
Panel B: Relationship between the Policy Exposure measure (EXP_j) and the baseline C-section rate in SUS, among municipalities where policy was binding



Notes: Panel A represents the distribution of the municipality-level C-section rate in SUS during the 12-month period prior to the policy announcement (i.e., baseline period), separately for two groups of municipalities. The first group (depicted in grey) consists of 1,495 municipalities where the policy was deemed non-binding, characterized by hospitals with baseline C-section rates below the 40% threshold. The second group (depicted in red) includes 1,265 municipalities where at least one hospital had a baseline C-section rates above 40%. Each point in Panel B represents an individual municipality in the latter group - those where the policy was deemed binding. The vertical axis illustrates our policy exposure measure, constructed in accordance with Equation (4.1), while the horizontal axis depicts the municipality's baseline C-section rate in SUS. To facilitate visualization, the plot excludes the top 1 percentile of the policy exposure measure.

Next, we present evidence confirming that the reimbursement cap operated as intended. To assess this evidence, we compare the hospitals' share of C-sections among reimbursed deliveries in SUS in the 12 months after the policy introduction relative to the analogous baseline shares. Figure 4.2 plots the distribution of these differences, at the hospital level. Hospitals with a baseline share above the 40% threshold experienced a decrease in their share of C-sections among reimbursed births, as indicated by the distribution of values to the left of zero. Conversely, hospitals with baseline rates of C-section below the 40% cap showed changes in the share of C-sections distributed around zero. This observation provides reassurance that the reimbursement cap effectively constrained hospitals that exceeded the introduced threshold. It is important to note, however, that a reduction in the share of reimbursed C-sections in SUS may not necessarily indicate an actual decrease in the relative frequency of this type of delivery. This is because uncompensated C-sections are excluded from SUS claims records.

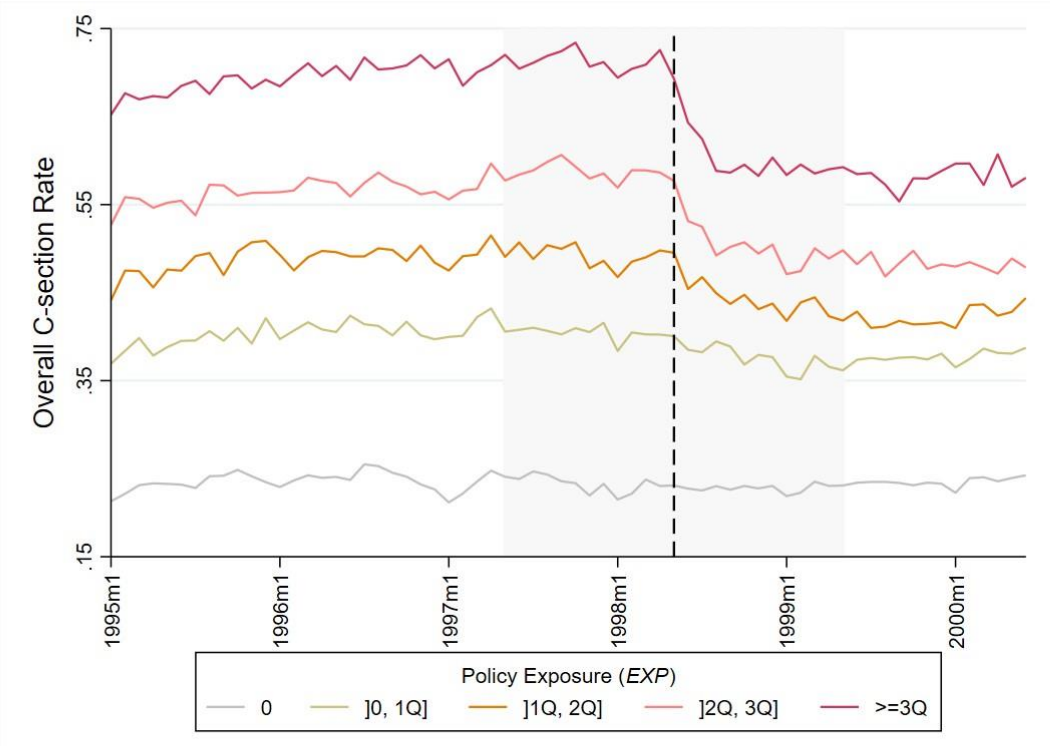
Figure 4.2: Differences between Post- vs Pre-policy share of C-sections among reimbursed deliveries in SUS hospitals



Notes: Shares of C-sections among reimbursed deliveries in each SUS hospital are computed for the 12-month periods before and after the policy. The figure presents the distribution of the difference in C-section shares between the post- and pre-policy periods, at the hospital level, separately for the groups of hospitals with SUS pre-policy C-section rate above and below (or equal to) 40%.

To evaluate the impact of the policy on C-section usage comprehensively, our analysis will consider all births registered in the national birth certificates. Figure 4.3 plots how the overall share of C-sections evolve over time for different classes of municipalities, according to our measure of policy exposure. The shaded area in the graph represents the period between the 12 months before the policy announcement until the 12 months after it, our analysis period. The dashed line denotes the month the threshold was announced. We observe that decreases in C-section rate following the policy announcement are restricted to the group of municipalities where the threshold was expectedly binding. Moreover, we observe that the magnitude of the decrease in C-section rates is positively associated with the level of exposure to the SUS threshold policy. To assess the policy's impacts, our empirical strategy exploits variation in this measure of (ex-ante) policy exposure across municipalities as a proxy for treatment intensity.

Figure 4.3: Evolution of C-section rate by municipality's exposure to the threshold policy



Notes: The plot presents the evolution of C-section rates among all live births registered in the country between January 1995 and May 2000 (last month before policy decentralization). Specifically, it displays the average rates of municipalities categorized into different groups based on their exposure to the introduced threshold. In grey, the plot shows the 1,495 municipalities where baseline rates suggest that the policy was not binding. The remaining municipalities are represented using varying shades of red, with darker shades indicating higher levels of exposure to the policy. The first group includes 317 municipalities in the first quartile of the distribution, while the subsequent quartiles have 316 municipalities each. Our policy exposure measure was constructed according to Equation (4.1), and the exact values of each quartile of the distribution can be found in Figure 4.1 (Panel B). During the years of 1995 and 1996, no information on month of occurrence was reported for 23% births. These are uniformly distributed over the months of the year for the given municipality where they were delivered. The dashed vertical line depicts the month of policy announcement (May, 1998), and the shaded area highlights the period between the 12th month before until the 12th month after the announcement (i.e., our analysis period).

Table 4.1 presents statistics on the characteristics of all live births in the country during the 12-month periods before and after the policy announcement (i.e., the shaded area in the previous plot). It compares post- vs pre-policy means among all municipalities and by groups of municipalities according to their assigned measure of policy exposure. As it was visually inspected in Figure 4.3, Table 4.1 Table 1 documents that the drop in the overall rate of C-sections is driven by municipalities where the policy was expectedly binding for SUS hospitals. There seems to be no important differences in the characteristics of mothers, pregnancies, and births over time, and the small observed changes are largely similar across the two municipality groups. In terms of outcomes at birth, we observe largely stable Apgar scores and a slight decrease in birthweight. Table 4.1 also indicates that the municipalities where the threshold policy was considered binding were generally larger, with approximately twice the number of deliveries compared to municipalities where the policy was deemed non-binding. Sample means of SUS hospitalization outcomes are provided in the relevant regression tables.

Table 4.1: Descriptive statistics: Pre- *vs* Post-policy

	All municipalities		Binding municipalities (EXP_j>0)		Non-binding municipalities (EXP_j=0)	
	Pre-Policy	Pos-Policy	Pre-Policy	Pos-Policy	Pre-Policy	Pos-Policy
Total N. municipalities	2,760	2,760	1,265	1,265	1,495	1,495
Total N. births	2,960,085	3,015,944	1,883,618	1,910,951	1,076,467	1,104,993
Municipality level						
N. births/municipality	1,072	1,093	1,489	1,511	720	739
C-section rate	37%	34%	55%	48%	23%	23%
Birth level						
C-section (%)	40.9	38.7	48.1	44.9	28.3	27.9
Age of mother	24.6	24.6	24.8	24.8	24.4	24.4
First birth (%)	0.31	0.33	0.34	0.35	0.27	0.29
Multiple pregnancy (%)	1.85	1.84	1.91	1.92	1.73	1.69
Birth in municipality of residence (%)	79.3	79.7	79.5	79.6	79.0	79.7
N. years of education: ≥8 (%)	36.2	36.5	41.0	41.6	27.6	27.9
N. prenatal visits: 0 (%)	6.98	6.05	5.17	4.47	10.21	8.80
N. prenatal visits: 1-6 (%)	42.5	43.6	38.3	39.4	50.0	51.0
N. prenatal visits: ≥7 (%)	50.5	50.3	56.5	56.1	39.8	40.2
N. weeks of gestation: <27 (%)	0.56	0.56	0.53	0.53	0.62	0.61
N. weeks of gestation: 28-36 (%)	4.82	5.24	5.16	5.62	4.23	4.60
N. weeks of gestation: 37-41 (%)	93.2	90.9	93.0	91.0	93.6	90.8
N. weeks of gestation: ≥42 (%)	1.40	3.25	1.32	2.82	1.54	4.01
Birthweight (g)	3,188	3,184	3,168	3,165	3,224	3,216
Birthweight < 2,500g (%)	7.93	7.99	8.39	8.47	7.11	7.16
Apgar score at 1st minute	8.05	8.05	8.07	8.08	8.01	8.00
Apgar score at 5th minute	9.13	9.15	9.16	9.19	9.07	9.09
5th min Apgar score < 7 (%)	2.04	1.92	1.90	1.77	2.32	2.20

Notes: Descriptive statistics were extracted from birth certificates containing all registered live births in the country. Statistics are reported, separately, for births delivered during the 12 months before the policy announcement (May 29, 1997, until May 28, 1998) and those delivered during the following 12 months (May 29, 1998, until May 28, 1999). Observations include all births during these time periods which occurred in municipalities with at least one registered birth during each one of the two time periods and at least one C-section performed in SUS during the pre-policy period. The last four columns of the table categorize the sample of municipalities based on their exposure to the introduced threshold. Our policy exposure measure was constructed following Equation (4.1). Non-binding municipalities are those assigned a zero-exposure measure (i.e., those where no hospitals presented a baseline SUS C-section rate above the threshold). Binding municipalities are those assigned a positive exposure measure (i.e., those where at least one hospital within the municipality exhibited a baseline SUS C-section rate exceeding the threshold).

4.5 Results

4.5.1 C-section likelihood and health at birth

Table 4.2 reports the effect of the policy on the likelihood of C-section. The first column includes the entire sample of 5,976,029 observations, described in Table 4.1. The estimation sample reduces as we add control variables. In the last column the estimation sample for the model with the largest number of covariates is used to estimate a model without them. This helps us disentangle changes in coefficients due to compositional differences between estimations from changes driven by conditioning results on covariates.

The estimated coefficient for the interaction term indicates a positive relationship between the level of exposure to the threshold policy and the magnitude of the decrease in the likelihood of C-section within the municipality. Results are robust to adding fixed effects of the exact year and month of birth and controlling for covariates.

Specifically, estimates suggest that the policy, on average, reduces the likelihood of C-section by 0.25 percentage points (p.p.) for every 1 p.p. increase in the municipality's baseline proportion of C-sections that would have gone uncompensated if the policy had been in place during that period. Notably, an increase of one standard deviation in our measure of policy exposure, equivalent to approximately 16 p.p., leads to a decrease in the likelihood of C-section of roughly 4 p.p. This reduction represents 10% of the average baseline outcome in our sample.[106]

Table 4.3 presents the main coefficients of interest from regressions that evaluate health outcomes at birth. Given the assessment of multiple different outcome variables, the table presents, in brackets, Romano-Wolf p-values for the interaction term coefficient. Because these p-values correct for multiple hypothesis testing, they offer a more cautious interpretation of the statistical significance of our coefficient of interest.

Consistent with the finding of post-policy drop in the choice for C-section, which is typically performed before the end of natural labor, the policy is shown to have caused birthweight to increase. While we also observe a decrease in the probability of low birthweight (i.e., below 2,500g), the effects are small and are accompanied by no statistically

[106] This is computed as -0.25*0.16/0.409, where the denominator corresponds to the share of births by C-section in the baseline period (see Table 4.1).

significant effects on probability of pre-term birth (i.e., before 37 weeks of pregnancy). We also find no significant effects on Apgar score measurements.

It's worth noting that, since our sample includes births delivered in the private sector to evaluate a policy directly targeting public hospitals, the results are expected to be not only robust to within-municipality spillovers but also conservative in terms of the magnitude of coefficients. Regarding this last point, our estimates are expected to be offset by the (presumable) absence of effects among private sector births occurring in municipalities with high baseline SUS C-section rates which would have happened in private hospitals irrespective of the policy. For this reason, our estimates should be interpreted as lower bounds.

Table 4.2: Policy effects on C-section likelihood

	(1) Baseline	(2) Add Year/Month FE	(3) Add Controls	(4) More Controls	(5) Baseline (sample (4))
	C-section likelihood				
$EXP * Pos$	-0.248***	-0.249***	-0.250***	-0.225***	-0.240***
	(0.014)	(0.014)	(0.014)	(0.017)	(0.016)
Pos	-0.001	-0.029***	-0.026***	-0.026***	0.002
	(0.002)	(0.004)	(0.004)	(0.004)	(0.002)
Mother's age			0.011***	0.011***	
			(0.001)	(0.000)	
Dummy for multiple pregnancy			0.170***	0.174***	
			(0.008)	(0.008)	
Dummy for higher-order birth				-0.060***	
				(0.004)	
Dummy for ≥ 8 years of education				0.146***	
				(0.004)	
Dummy for 1-6 prenatal visits (vs 0)				0.093***	
				(0.004)	
Dummy for 7+ prenatal visits (vs 0)				0.212***	
				(0.005)	
N. observations	5,976,029	5,976,029	5,856,396	3,923,004	3,923,004
Municipality FE	X	X	X	X	X
Year/Month FE		X	X	X	

Notes: Observations refer to individual-level (live) births that took place in the period between the 12 months before and the 12 months after the policy announcement on May 29, 1998. Births include all childbirth deliveries presented in Table 4.1. Results are based on the estimation of Equation (4.2). The outcome variable is the likelihood of C-section delivery. In Model (1), we regress the outcome on the two first variables described in the first column, an intercept, and municipality fixed effects. Model (2) adds year/month fixed effects. Models (3) and (4) account for the control variables detailed in the table's first column. Finally, Model (5) estimates the first model (no control variables) using the Model (4)'s estimated sample. Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

Table 4.3: Policy effects on health outcomes at birth

	(1) Baseline	(2) Add Year/Month FE	(3) Add Controls	(4) More Controls	(5) Baseline (sample (4))
			Birthweight		
$EXP * Pos$	49.565***	49.401***	48.925***	60.307***	50.952***
	(8.301)	(8.290)	(8.217)	(8.051)	(7.604)
	[0.004]	[0.004]	[0.004]	[0.0040]	[0.0040]
Pos	-6.910***	-12.096***	-10.366***	-9.795**	-8.774***
	(1.147)	(3.891)	(3.829)	(4.035)	(1.335)
N. obs.	5,932,428	5,932,428	5,817,384	3,904,119	3,904,119
			P(Birthweight < 2,500g)		
$EXP * Pos$	-0.0067**	-0.0067**	-0.0066***	-0.0097***	-0.0058**
	(0.0026)	(0.0026)	(0.0025)	(0.0030)	(0.0030)
	[0.012]	[0.008]	[0.008]	[0.004]	[0.036]
Pos	0.0009**	0.0023	0.0020	0.0004	0.0010*
	(0.0004)	(0.0022)	(0.0022)	(0.0024)	(0.0005)
N. obs.	5,932,428	5,932,428	5,817,384	3,904,119	3,904,119
			P(Gestational age < 37 weeks)		
$EXP * Pos$	-0.004	-0.003	-0.003	-0.001	0.000
	(0.009)	(0.009)	(0.009)	(0.006)	(0.006)
	[0.857]	[0.884]	[0.868]	[0.940]	[0.948]
Pos	0.004*	0.002	0.002	-0.000	0.003**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
N. obs.	5,840,905	5,840,905	5,735,072	3,885,711	3,885,711
			Apgar score at 1st minute		
$EXP * Pos$	0.048	0.048	0.050	0.041	0.029
	(0.036)	(0.036)	(0.036)	(0.040)	(0.040)
	[0.211]	[0.215]	[0.195]	[0.442]	[0.805]
Pos	0.005	0.001	0.000	0.008	0.010
	(0.007)	(0.011)	(0.011)	(0.012)	(0.007)
N. obs.	5,153,556	5,153,556	5,074,559	3,488,192	3,488,192
			Apgar score at 5th minute		
$EXP * Pos$	0.055	0.055	0.055	0.040	0.028
	(0.035)	(0.035)	(0.035)	(0.040)	(0.040)
	[0.123]	[0.135]	[0.088]	[0.442]	[0.805]
Pos	0.017***	-0.002	-0.003	0.007	0.020***
	(0.005)	(0.009)	(0.009)	(0.010)	(0.006)
N. obs.	5,090,378	5,090,378	5,016,128	3,455,004	3,455,004
			P(5th min. Apgar score < 7)		
$EXP * Pos$	-0.0002	-0.0003	-0.0006	-0.0013	-0.0005
	(0.0027)	(0.0027)	(0.0027)	(0.0028)	(0.0028)
	[0.936]	[0.884]	[0.868]	[0.753]	[0.948]
Pos	-0.0017***	-0.0001	-0.0003	-0.0010	-0.0014***
	(0.0004)	(0.0011)	(0.0011)	(0.0012)	(0.0004)
N. obs.	5,090,378	5,090,378	5,016,128	3,455,004	3,455,004
Municipality FE	X	X	X	X	X
Year/Month FE		X	X	X	
Controls			X	X	
More Controls				X	

Notes: Observations refer to individual-level (live) births that took place in the period between the 12 months before and the 12 months after the policy announcement on May 29, 1998. Births include all childbirth deliveries presented in Table 4.1. Results are based on the estimation of Equation (4.2). Outcomes are described in the body of the table. The estimated models are the same as those from Table 4.2. Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1). In brackets, we present Romano-Wolf p-values for the interaction term coefficient. These p-values were obtained following the procedure outlined by Clarke (2021), using 250 bootstrap replications.

4.5.2 Foetal, infant, and maternal mortality

Next, we evaluate whether the policy affected mortality. We look at the number of stillbirths (i.e., fetal mortality), infant deaths (i.e., during their first year of life), and postpartum maternal deaths as outcome variables. Table 4.4 presents the results. The estimated coefficient for the interaction term is not statistically significantly different from zero for any of these outcomes. In the Appendix, Table C.2 presents results for the subsample of municipalities with at least one registered death. It also includes results from log-transformed models, which consider this same subsample given that the natural logarithm function is not defined at zero. Log-transformed models estimated in these subsamples reveal negative effects on fetal and infant deaths that are statistically significant at conventional levels, after accounting for multiple hypothesis testing.

Table 4.4: Policy effects on mortality

	Number of deaths		
	Stillbirths	Infant	Maternal
$EXP * Pos$	-0.876	0.656	-0.105
	(0.846)	(1.109)	(0.082)
	[0.247]	[0.355]	[0.175]
Pos	0.431	-0.963	-0.002
	(0.452)	(0.628)	(0.027)
N. obs.	5,520	5,520	5,520
Municipality FE	X	X	X
Baseline mean	*13.99*	*23.71*	*0.489*

Notes: Observations from death certificates are aggregated at the municipality-time level. The table presents results based on the estimation of Equation (4.3), where the outcome corresponds to the number of registered deaths of fetuses, infants, and mothers. Infant mortality refers to the death of children within their first year of life, while maternal mortality relates to the death of women due to reasons associated with childbirth. For stillbirths and infant deaths, time refers to childbirth deliveries within the 12-month periods before or after the policy announcement on May 29, 1998. For maternal mortality, time represents deaths occurring during the puerperium within the 12-month periods before or after the policy announcement. Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1). In brackets, we present Romano-Wolf p-values for the interaction term coefficient. These p-values were obtained following the procedure outlined by Clarke (2021), using 250 bootstrap replications.

4.5.3 Infant and maternal hospitalizations in SUS

Table 4.5 presents the policy effects on number of infant admissions in the public sector during their first year of life. We observe a decrease in the total number of hospitalizations of infants born during the 12 months following the policy introduction relative to those born in the preceding 12 months that is specific to the extent to which the threshold policy was binding to municipalities. Estimates suggest that an increase of one standard deviation in the municipality's measure of policy exposure (equivalent to ~16 p.p.) would result in a fall in the number of hospitalizations of 0.16*(18.639) = 3.03, which is equivalent to 1.3% of the average number of admissions of children born during the pre-policy period.

The following columns of the table inform that the decrease is driven by admissions of children very early in life. A one standard deviation increase in the municipality's level of exposure to the policy is associated with a decrease of roughly 2.5% in the number of hospitalizations of children during the first quarter after they are born. During the second quarter, the decrease represents 1.5% of the baseline average. Effects estimated for the second semester of life are not statistically different from zero.

Estimates obtained from log-transformed models, as presented in Table C.3, tend to exhibit larger magnitudes and more frequent statistical significance.[107] Given the sample selection resulting from these restrictions and the fact that the estimates from log-transformed models tend to be generally stronger (both in terms of magnitude and statistical significance), considering outcomes measured as counts is deemed a more conservative approach, making it our preferred outcome specification.

Table 4.6 investigates results by hospitalization type. We find a significant decrease in the number of hospitalizations from chronic pulmonary disorders, such as asthma and bronchitis. Relative to the sample mean, a one standard deviation rise in the municipality's level of policy exposure triggers a fall of 3.5% in the number of infant admissions due to these respiratory problems. No impacts are found for infant admission to the hospitals' intensive care unit (ICU). This is consistent with the literature on the health consequences of potentially avoidable C-sections, which typically reports infant respiratory symptoms (Card et al., 2023; Costa-Ramón et al., 2021; Jachetta, 2016) but no severe health impact such as mortality and admission to the intensive care unit (Amaral-Garcia et al., 2022; Costa-Ramón et al., 2018). Finally, the last column reports no effects on maternal hospitalizations due to post-partum complications.

[107] Coefficients of log-level regression models are measured in percentage terms. Take the estimate on total number of infant hospitalizations of -0.261. It infers a decrease of 0.261% in the number of infant hospitalizations following a one percentage point increase in the level of policy exposure faced by the municipality. A 16 percentage point increase in the latter (equivalent to 1 standard deviation) would lead to a 4.2% increase in the evaluated outcome.

Table 4.5: Policy effects on SUS hospitalizations during first year of life

	Number of Infant Hospitalizations (≤1yo)				
		Age at admission			
	All	0-3 months	3-6 months	6-9 months	9-12 months
$EXP * Pos$	-18.639**	-13.701***	-4.571**	-0.738	0.371
	(8.068)	(4.386)	(2.056)	(2.002)	(2.060)
	[0.008]	[0.004]	[0.008]	[0.940]	[0.940]
Pos	4.279*	6.881***	-0.494	-0.852	-1.256*
	(2.400)	(1.572)	(0.541)	(0.602)	(0.670)
N. obs.	5,520	5,520	5,520	5,520	5,520
Municipality FE	X	X	X	X	X
Baseline mean	*227.3*	*86.56*	*48.03*	*46.92*	*45.77*

Notes: Observations from SUS hospital claims are aggregated at the municipality-time level, where time refers to indicators of patients' date of birth during the 12-month periods before or after the policy announcement on May 29, 1998. The table presents results based on the estimation of Equation (4.3), where the outcome corresponds to the number of infant hospitalizations (i.e., admissions up to 365 days old). The column titles describe the time horizon after birth for which the outcome is evaluated. Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1). In brackets, we present Romano-Wolf p-values for the interaction term coefficient. These p-values were obtained following the procedure outlined by Clarke (2021), using 250 bootstrap replications.

Table 4.6: Policy effects on SUS hospitalizations, by type

	Number of Infant Hospitalizations (≤1yo)			Number of Maternal Hospitalizations
	All	ICU	Pulmonary disorders	Post-partum complication
$EXP * Pos$	-18.639**	0.454	-1.940**	0.011
	(8.068)	(1.493)	(0.861)	(0.021)
	[0.008]	[0.940]	[0.008]	[0.940]
Pos	4.279*	0.571	1.098***	-0.010*
	(2.400)	(0.530)	(0.250)	(0.005)
N. obs.	5,520	5,520	5,520	5,520
Municipality FE	X	X	X	X
Baseline mean	*227.3*	18.63	8.804	0.0143

Notes: Observations from SUS hospital claims are aggregated at the municipality-time level. The table presents results based on the estimation of Equation (4.3) for outcomes specified in the title of the columns. Hospitalizations were classified based on their primary diagnoses, which were recorded using the ICD-9 system prior to 1998 and the ICD-10 system thereafter. Chronic pulmonary diseases refer to ICD codes 490-496 (ICD-9) & J40-J47 (ICD-10), while Postpartum complications correspond to 670-677 (ICD-9) & O85-O92 (ICD-10). Hospitalizations in the Intensive Care Unit (ICU) are classified as episodes where the patient spent at least one day in the ICU of the hospital. For infant hospitalizations (i.e., admissions up to 365 days old), time refers to date of birth within the 12-month periods before or after the policy announcement on May 29, 1998. For maternal disorders, time represents post-partum complications occurring within the 12-month periods before or after the policy announcement. Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1). In brackets, we present Romano-Wolf p-values for the interaction term coefficient. These p-values were obtained following the procedure outlined by Clarke (2021), using 250 bootstrap replications.

Our analysis of health outcomes after birth relies on the assumption that infants are born in the same municipality as they are admitted to the hospital. Although our hospital claims data lack information on infants' municipality of birth, it includes their municipality of residence. Notably, the proportion of infants hospitalized within their municipality of residence closely mirrors the proportion of mothers giving birth in the same municipality where they reside, suggesting similar patterns of movement away from the municipality of residence during childbirth and infant hospitalization.[108] We also investigate whether the policy induced changes in the proportions of infant hospitalizations and childbirth deliveries from patients traveling from other municipalities. Table C.4 in the appendix shows no clear association between these proportions and the extent to which the threshold was binding in the given municipality. In other words, it indicates that mothers do not change decisions, in response to the policy, regarding which municipalities to give birth and admit sick children.

A limitation of our analysis of hospitalization outcomes is that our data is restricted to admissions to SUS hospitals. The validity of our results could be questioned if mothers seeking care in municipalities more exposed to the policy became more likely to admit their children to private hospitals after the introduction of the C-section threshold policy.[109] In such a scenario, our findings of decreases in SUS hospitalizations could simply reflect hospitalizations that would have happened in SUS occurring in the private sector instead. To address this concern, we analyze SUS hospitalizations for the same cohort of children, focusing on diagnoses largely independent of the event of childbirth. The absence of effects for these placebo outcomes, discussed in Section 4.7.1, provides strong evidence against the possibility of a shift of infant hospitalizations to private hospitals. Therefore, it reinforces the understanding that the estimated drop in infant hospitalizations, attributed to causes previously linked to C-sections,[110] was indeed a consequence of the decline in C-section likelihood induced by the policy.

Another hypothesis is that the fall in hospitalizations of infants born after the policy introduction is driven by changes in the overall quality experienced during childbirth. This

---

[108] Episode-level data on hospitalizations of infants born during our analysis' time horizon indicates that 78.7% of admissions took place in the patient's municipality of residence. Births certificates report that 79% of all births in our sample occurred in the mothers' municipality of residence (see Table 4.1). According to the considerations outlined in footnote 102, mothers who give birth outside their municipality of residence are likely to travel to the same municipality when hospitalizing sick children.

[109] This could happen if, for instance, mothers became more likely to give birth in private facilities (as a response to the policy) and consequently decided to admit their children in the same facilities.

[110] Costa-Ramón et al. (2021) and Jachetta (2016) find increased infant inpatient admissions and outpatient visits due to respiratory problems. Card et al. (2023) show that infants quasi-randomly delivered at hospitals with higher propensity to perform C-sections are more likely to visit the emergency service for respiratory-related issues.

scenario could arise if mothers became more likely to deliver in the private sector to ensure a C-section, where the quality of care is generally higher. Our findings effectively dismiss this notion. First, if women seeking C-sections consistently shifted to the private sector, our analysis would have indicated minimal impacts on C-section likelihood. Table 4.2 reports a substantial decrease in C-section use. Event study analysis, presented later on, further reinforces this finding by demonstrating that the decrease occurred promptly after the threshold implementation and persisted over time. Second, if the decline in infant hospitalizations resulted from overall improvements in providers' quality rather than a shift in delivery methods, we would anticipate improvements in birth outcomes such as Apgar scores. Instead, we observe effects solely on birthweight (see Table 4.3), an outcome directly influenced by the timing of birth, which is mechanically linked to the mode of childbirth delivery (i.e., C-sections allow for births before labor onset, leading to lower birthweights). All the above evidence supports the interpretation that our findings stem from the policy-induced shift in delivery methods.

4.6 Policy effect heterogeneity

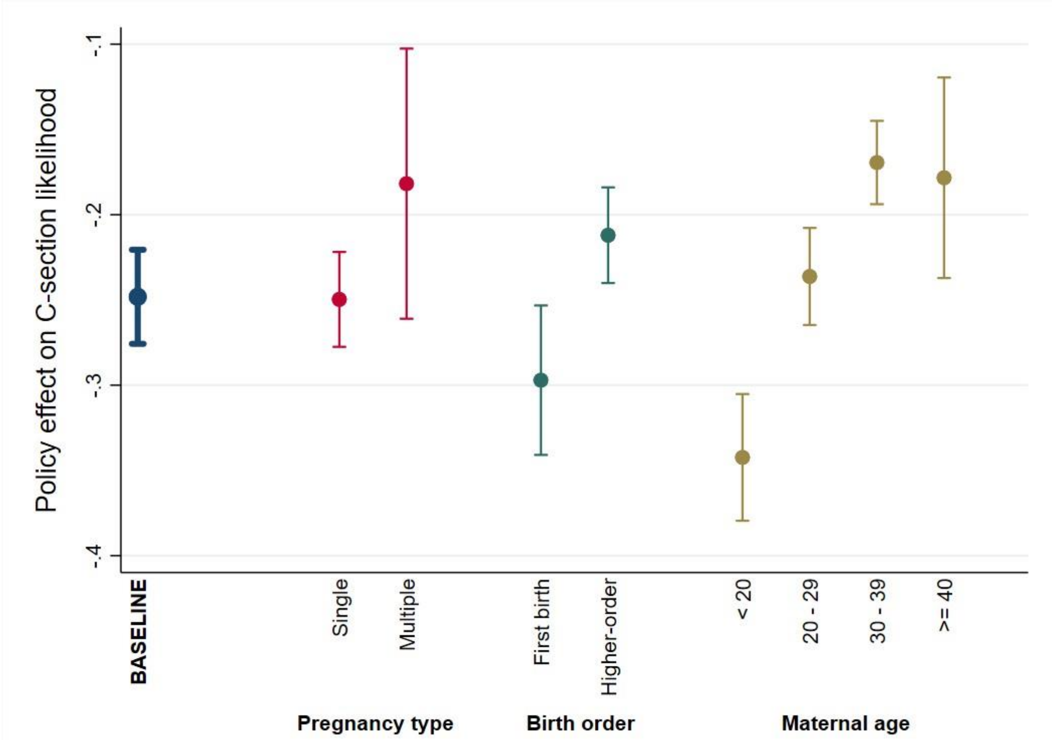
This section looks at heterogeneity in the extent to which the policy caused C-section likelihood to decrease. First, we report heterogeneous results by ex-ante characteristics of births. Next, we estimate a more flexible model that allows for marginal effects to vary for different levels of policy exposure.

4.6.1 By risk profile of birth

This section sheds light on the heterogeneous effects for different types of births which are more *vs.* less likely to receive medical recommendation for C-section. Three ex-ante characteristics are considered: pregnancy type (single *vs* multiple), birth type (first *vs* higher-order birth), and maternal age. Births from multiple pregnancies and from older mothers are generally associated with higher risks. Given that prior C-sections typically constitute a medical indication for repeated C-section, higher-order births have increased chances of receiving indication of C-section delivery.

Figure 4.4 documents higher decreases among deliveries from single pregnancies (relative to multiple pregnancies), from first births (relative to higher-order deliveries), and from younger (relative to older) mothers. This is consistent with the understanding that C-section decreases brought by the policy came especially from medically unjustified C-sections, which were commonly performed prior to the policy reform.

Figure 4.4: Heterogeneous effects by type of birth



Notes: The figure presents the 95% confidence interval of policy effects estimates. It includes the baseline results from Table 4.2 (Model 1) and shows heterogeneous results based on three characteristics: pregnancy type, birth order, and maternal age. To estimate heterogeneous effects separately in each of these margins, we incorporate interaction terms in Equation (4.2) between the characteristic indicator (e.g., multiple birth) and three variables: (i) post-policy dummy, (ii) municipality's measure of policy exposure, and (iii) the interaction between the previous two terms. The coefficient of this triple interaction indicates the incremental effects of the policy (for each percentual point increase in our measure of policy exposure) for these respective types of births. Standard errors are clustered at the municipality level. The data used in the regressions consist of individual-level observations encompassing all live births in the country that took place in the period between the 12 months before and the 12 months after the policy announcement on May 29, 1998.

4.6.2 By different levels of policy exposure

The results presented so far leverage variation in treatment intensity resulting from the introduction of a fixed reimbursement cap to the relative use of C-sections across hospitals with different baseline propensities for this type of delivery. Our empirical strategy assumes that policy effects are linear in our measure of treatment intensity. According to our main specification, a percentual point increase in the municipality's measure of policy exposure is considered to affect the post-policy C-section likelihood in similar ways across all exposure levels (i.e., constant marginal effects).

In this section, we estimate a more flexible model that does not impose any parametric relationship between our measure of policy exposure and the outcome variable. The continuous measure of threshold exposure (EXP_j) in Equation (4.2) is replaced with a set of four indicator variables, denoted as $1_{EXP_j^+ \ in \ Qk}$, where k corresponds to each of the four quartiles of the distribution of policy exposure illustrated in Figure 4.1, Panel B. The specific regression model for estimation is provided below.

$$CS_{ijym} = \alpha + \sum_{k=1}^{4} \boldsymbol{\beta_k} \left\{ 1_{EXP_j^+ \, in \, Qk} * Pos_{ym} \right\} + Pos_{ym} + \mu_j + \varepsilon_{ijym} \qquad \textbf{(4.4)}$$

where $1_{EXP_j^+ \, in \, Qk}$ takes value 1 if municipality *j*'s measure of threshold exposure falls within quartile *k*, and 0 otherwise. The omitted category refers to municipalities with a threshold exposure measure equal to 0. The parameters β_1, β_2, β_3, and β_4 capture differences in the average change in C-section likelihood following the policy introduction among municipalities in different quartiles of the constructed measure of threshold exposure relative to the municipalities where the policy was not binding.

To compare changes in C-section likelihood among all municipalities where the introduced threshold was binding to those where it was non-binding, a simpler model was estimated where the indicators for each quartile were replaced by a single indicator variable that takes the value 1 for positive values of exposure and 0 otherwise.

Results from these two specifications are presented in Table 4.7, along with the baseline results reported earlier. The estimates demonstrate that the policy effects are statistically significant across all specifications, regardless of how the exposure measure is incorporated in the estimation model. The middle column reveals a systematic decrease in C-section likelihood in municipalities where the threshold was binding. The estimation presented in the last column reinforces the message highlighted by our baseline results, without imposing any parametric relationship between the estimated effects and our constructed measure of exposure to the introduced threshold. It demonstrates that the effect's magnitude and statistical significance increase with the extent to which the introduced threshold was expected to be binding.

Table 4.7: Policy effects on C-section likelihood, using different specifications

	C-section likelihood		
	Linear in exposure (Baseline)	Any exposure	By exposure quartile
$EXP * Pos$	-0.248***		
	(0.014)		
$1_{EXP^+} * Pos$		-0.029***	
		(0.004)	
$1_{EXP^+ \; in \; Q1} * Pos$			-0.006**
			(0.003)
$1_{EXP^+ \; in \; Q2} * Pos$			-0.029***
			(0.007)
$1_{EXP^+ \; in \; Q3} * Pos$			-0.069***
			(0.006)
$1_{EXP^+ \; in \; Q4} * Pos$			-0.115***
			(0.009)
Pos	-0.001	-0.003*	-0.003*
	(0.002)	(0.001)	(0.001)
N. obs.	5,976,029	5,976,029	5,976,029
Municipality FE	X	X	X

Notes: Observations refer to individual-level (live) births that took place in the period between the 12 months before and the 12 months after the policy announcement on May 29, 1998. Births include all childbirth deliveries presented in Table 4.1. The outcome variable is the likelihood of C-section delivery. The first estimation replicates results presented in the first column of Table 4.2. The second estimation replaces the continuous values of the policy exposure measure with an indicator variable of positive values of exposure (i.e., EXP>0). The third estimation considers, instead, indicator variables for each quartile of the distribution of the exposure measure among municipalities where the threshold was binding. The corresponding model is described by Equation (4.4). The omitted category of all estimations refers to municipalities where the introduced threshold was deemed non-binding (i.e., EXP = 0). Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

Table 4.8 reassess the policy impacts on the number of infant hospitalizations in SUS using these same specifications. The final column indicates that the decline is primarily observed across municipalities in the top quartile of exposure to the policy. In these municipalities, at least 37% of C-sections performed in the baseline period would not have been compensated if the reimbursement cap had been in place. The average baseline C-section rate of municipalities in this top quartile was 70%, a rate that appears implausibly high from a medical standpoint.

Table 4.8: Effect on number of infant hospitalizations in SUS, by specification

	All infant hospitalizations			Pulmonary disorders		
	Linear in exposure (Baseline)	Any exposure	By exposure quartile	Linear in exposure (Baseline)	Any exposure	By exposure quartile
$EXP * Pos$	-18.639**			-1.940**		
	(8.068)			(0.861)		
$1_{EXP^{+}} * Pos$		0.368			0.479	
		(3.980)			(0.428)	
$1_{EXP^{+}\ in\ Q1} * Pos$			8.406			2.165*
			(10.543)			(1.211)
$1_{EXP^{+}\ in\ Q2} * Pos$			1.854			0.284
			(7.096)			(0.650)
$1_{EXP^{+}\ in\ Q3} * Pos$			-1.042			0.408
			(5.992)			(0.649)
$1_{EXP^{+}\ in\ Q4} * Pos$			-7.773***			-0.947***
			(2.999)			(0.330)
Pos	4.279*	2.061	2.061		0.665***	0.665***
	(2.400)	(1.965)	(1.965)		(0.202)	(0.202)
N. obs.	5,520	5,520	5,520	5,520	5,520	5,520
Municipality FE	X	X	X	X	X	X

Notes: Observations from SUS hospital claims are aggregated at the municipality-time level. Infant hospitalizations refer to admissions up to 365 days old within the 12-month periods before or after the policy announcement on May 29, 1998. Chronic pulmonary diseases concern diagnoses informed by ICD codes 490-496 (ICD-9) & J40-J47 (ICD-10). For each outcome, the first estimation replicates results presented in Table 4.5 and Table 4.6. The second estimation replaces the continuous values of the policy exposure measure with an indicator variable of positive values of exposure (i.e., EXP>0). The third estimation considers, instead, indicator variables for each quartile of the distribution of the exposure measure among municipalities where the threshold was binding. The omitted category of all estimations refers to municipalities where the introduced threshold was deemed non-binding (i.e., EXP = 0). Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

4.7 Robustness analyses

The main threat to our analysis is self-selection of expectant mothers to municipalities less constrained by the threshold policy aiming at increasing their chances to give birth by C-section.

There are conceptual and institutional reasons why this is unlikely to have been the case. First, the policy was not advertised in the media given that it targeted public providers. Besides, information on hospitals' C-section rates is not publicly available. Mothers are, therefore, not expected to have enough knowledge to be able to react to the policy. Second, given that the threshold was set at the hospital level, informed mothers would likely prioritize moving to other local hospitals (if available) instead of travelling to another municipality. Third, and most importantly, SUS patients cannot freely choose which municipalities to seek care in. Access is restricted to their municipality of residence or to a predetermined outsourced municipality in case the required service is not locally available.[111] Given all of the above and the policy's short notice during a tumultuous period of life, a more plausible and economically sensible scenario would be that, rather than relocating to other municipalities to access different SUS providers, mothers opt to switch to the private sector within their current municipality of residence. Since our analysis on health outcomes at birth encompasses all deliveries within a given municipality, our results remain robust to private sector migration, as long as it doesn't involve a change in the municipality where mothers deliver their babies.[112]

While the institutional features discussed above strongly suggest against selective migration of mothers across municipalities, this section offers empirical support for this. We begin by presenting event study results indicating that C-section likelihood did not exhibit significant increases among municipalities less constrained by the threshold in the 12 months following its implementation.

Whereas our results on outcomes at birth are robust to private sector migration within the given municipality, analogous movements in the context of infant hospital admissions would invalidate the causal interpretation of our findings on hospitalization outcomes.[113] We conclude the session by investigating whether the reported falls in SUS hospitalizations could be a simple reflect from post-policy increases of admitting sick children to private hospitals among municipalities where the policy was expectedly more binding. Results from placebo tests contradict this alternative hypothesis.

4.7.1 Event study

---

[111] For more details, see Section 4.2.1.

[112] There are no reasons to expect that mothers who switch to the private sector also change the municipality where they seek care in, except in the very special circumstances where there are public facilities but no private hospitals or when there are private hospitals but no public facility offering childbirth services in the mothers' municipality of residence (in which case SUS would have required them to travel to another municipality to give birth).

[113] The reason for this is that while the data employed in our analysis of outcomes at birth include all deliveries in the country (hence, municipality-level results incorporate any within-municipality movements), our analysis on hospitalization outcomes is based on data solely from SUS hospital admissions. If infants born after the policy announcement become more likely to be admitted to private hospitals among municipalities more exposed to the reform, unobserved observations (i.e., admission to private hospitals) would not be distributed at random and, therefore, would confound our results on policy impacts.

This event study analysis provides robust evidence on pre-trends and longer-term dynamics among groups of municipalities more and less constrained by the threshold. It considers a time horizon that is twice as long as contemplated in our main results.

An indicator of C-section delivery is regressed on month indicator variables from 24 months before until 24 months after the policy announcement, along with municipality-specific effects, where the omitted category corresponds to the month prior to the announcement of the threshold. Robust standard errors are clustered at the municipality level. The estimation is carried out separately for the groups of municipalities where the policy was binding (assigned a positive exposure measure) and non-binding (assigned a zero-exposure value). Panel A of Figure 4.5 exhibits the 95% confidence intervals for the estimated month-by-month coefficients, centered around the announcement of the threshold policy (May 1998, indexed as month 0), marked by the dashed vertical line. The subsequent dashed lines correspond to months when the threshold was updated.[114] The gray shaded area corresponds to our main period of analysis (from 12 months before to 12 months after the announcement). In Panel B, a similar plot illustrates both the 95% confidence interval and point estimates for municipalities in each quartile of our constructed measure of policy exposure.

The event study results reveal a significant and immediate decrease in the likelihood of C-sections following the threshold introduction, particularly in municipalities where the initial threshold was considered binding based on pre-policy propensities for C-sections. Not only was the decline immediate, but it also persisted over time. Panel B shows that the decrease increases in magnitude as the municipality's assigned measure of exposure rises. In contrast, in municipalities where the threshold was anticipated to be non-binding, there is no discernible discontinuity in the relative frequency of C-sections at the time the policy was implemented.

While the post-policy dynamics among municipalities in the control group (i.e., those where the 40% threshold was considered non-binding) remain relatively stable during our main analysis period (shaded in gray), we observe an upward trend after the threshold updated. In Table C.1 in the appendix, we present results from estimations where we address the fact that, towards the end of our analysis period, the threshold was revised from 40% to

[114] As described in Section 4.2.2, the 40% threshold was announced in May 1998. It was adjusted to 37% in January 1999. One year later, in January 2000, it was further reduced to 35%.

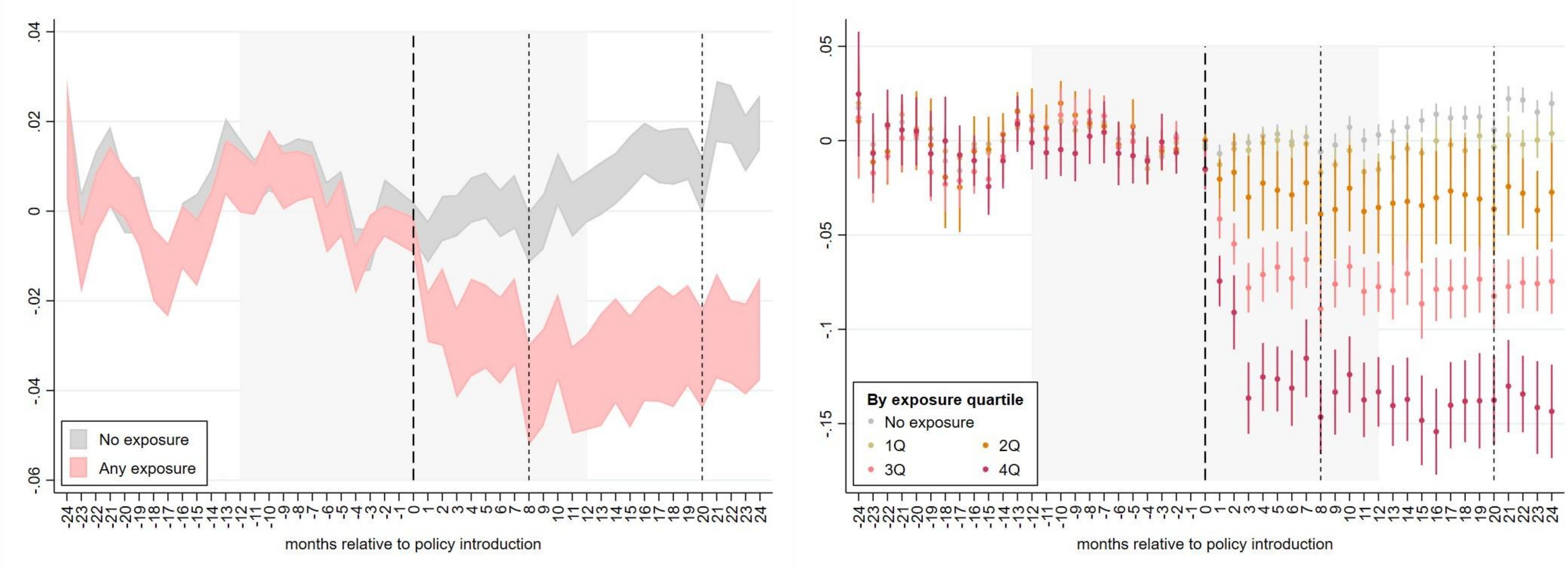
37%. These attempts involved excluding observations after the threshold update, introducing additional independent variables to model incremental effects following this update, as well as reconstructing our exposure measure based on the revised threshold level (37% instead of 40%). The results remain robust across all these alternative models.

Finally, we observe parallel trends in C-section likelihood during the months preceding the policy announcement in groups of municipalities facing different levels of exposure to the introduced threshold. These pre-policy trajectories are not only parallel but also very similar in levels. Overall, the evidence in this section is consistent with the identification assumption that post-policy trajectories of municipalities with lower exposure to the initial threshold serve as robust counterfactuals, in the absence of its introduction, to what would have occurred in municipalities with higher levels of exposure.

Figure 4.5: C-section likelihood during months around policy announcement

Panel A: Any (vs no) exposure to introduced threshold

Panel B: By exposure quartile to introduced threshold



Notes: The figures display the 95% confidence intervals of month-by-month coefficients obtained from a regression where the C-section indicator is regressed against month indicators and municipality fixed effects. The month of threshold announcement is indexed as month 0, marked by the first dashed vertical line. Subsequent dashed lines correspond to months when the threshold level was revised. The gray shaded area represents our main period of analysis (from 12 months before to 12 months after the announcement). Regressions are separately estimated for different groups of municipalities based on their exposure to the threshold introduction. Panel A depicts coefficients for municipalities where the policy was binding (positive exposure measure) and non-binding (zero-exposure value). Panel B compares coefficients between the latter and municipalities in each quartile of our constructed policy exposure measure. Robust standard errors are clustered at the municipality level. The data used in the regressions consist of individual-level observations encompassing all live births in the country that took place in the period between the 24 months before and the 24 months after the month of policy announcement (May 1998).

4.7.2 Placebo tests

While there is evidence suggesting that mothers admitting their children to SUS hospitals do not exhibit changes, as a response to the policy, in which municipality to seek care, they could still have altered their decisions regarding which hospital to admit their children within the given municipality. Our investigation on hospitalization outcomes is robust to compositional changes across SUS hospitals located in the same municipality; however, switching between the public and private sectors would threaten the causal interpretation of our results on hospitalization outcomes in SUS.

To investigate the plausibility of the alternative hypothesis that our findings on decreases in infant hospitalizations in the public sector are driven by post-policy-born children becoming more likely to be admitted to private hospitals among municipalities more constrained by the threshold policy, we study the policy effects on placebo outcomes measured by infant hospitalizations due to external causes. Because such diagnoses are, presumably, unrelated to the event of childbirth, we should find no impacts from a policy aimed at influencing decisions on method of childbirth delivery. Findings of negative coefficients for these placebo outcomes would give credit to the alternative hypothesis that decreases in SUS hospitalizations simply reflect corresponding increases in (unobserved) privately funded hospital admissions. Evidence of no impacts on these placebo hospitalizations would offer reassurance that the policy is indeed the driving force behind our findings of hospitalization declines.[115] Consistent with the latter scenario, Table 4.9 reports no policy effects on infant hospitalization due to external causes.

---

[115] Such interpretation of placebo results relies on the assumption that changes in mothers' decisions of whether to admit their children in SUS *vs* the private sector is not influenced by type of health problem. That is, if mothers became more likely to admit infants in private hospitals in the event of respiratory issues, this would also apply in case of conditions developed during pregnancy or in case of accidents.

Table 4.9: Policy effects on number of infant hospitalizations in SUS due to placebo diagnoses

	Infant hospitalizations due to External Causes		
	Linear in exposure	Any exposure	By exposure quartile
$EXP * Pos$	-0.117		
	(0.225)		
$1_{EXP^+} * Pos$		-0.017	
		(0.112)	
$1_{EXP^+\ in\ Q1} * Pos$			0.140
			(0.340)
$1_{EXP^+\ in\ Q2} * Pos$			-0.201
			(0.186)
$1_{EXP^+\ in\ Q3} * Pos$			0.043
			(0.126)
$1_{EXP^+\ in\ Q4} * Pos$			-0.049
			(0.097)
Pos	0.010***	0.012***	0.012***
	(0.002)	(0.003)	(0.003)
N. obs.	5,518	5,518	5,518
Municipality FE	X	X	X

Notes: Observations from SUS hospital claims are aggregated at the municipality-time level. Infant hospitalizations refer to admissions up to 365 days old within the 12-month periods before or after the policy announcement on May 29, 1998. External causes are categorized as diagnoses falling within chapters 17 or 19 (ICD-9 system) or chapters 19 or 20 (ICD-10 system). The first estimation is based on Equation (4.3). The second estimation replaces the continuous values of the policy exposure measure with an indicator variable of positive values of exposure (i.e., EXP>0). The third estimation considers, instead, indicator variables for each quartile of the distribution of the exposure measure among municipalities where the threshold was binding. The omitted category of all estimations refers to municipalities where the introduced threshold was deemed non-binding (i.e., EXP = 0). Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

4.7.3 Post-policy changes in amount reimbursed

Another advantage of conducting our analysis at the municipality level is that it corresponds to the level to which federal payments are transferred.[116] Based on information on federal reimbursements of childbirth procedures extracted from SUS hospital claims data, we inspect whether municipalities more exposed to the introduced threshold experienced changes in reimbursement revenues following its implementation. Estimations are, again, based on the model described by Equation (4.3). Reimbursed amounts are measured in prices of December 2022.

[116] Although the threshold was implemented at the hospital level, federal reimbursement payments are transferred to the local authority (i.e., typically the health department of the municipal government) managing the provision of care in hospitals within their catchment area. See Section 4.2.1 for more details.

Table 4.10 presents estimates from various models. The first model shows the average change in municipalities' revenues from reimbursed childbirth deliveries during the 12-month post-policy period compared to the 12 months preceding the policy introduction, indicating an average increase of 51% among all municipalities in our sample.[117] This aligns with the fact that, alongside the introduction of the reimbursement cap affecting the relative use of C-sections, the reform increased the unit fees for all reimbursed deliveries.

The subsequent estimations incorporate measures of exposure to the threshold policy as explanatory variables, enabling us to distinguish the post-policy changes between municipalities that faced binding thresholds and those that did not. Model 2 adds our continuous policy exposure measure and shows a negative linear relationship between post-policy changes in the total reimbursed amount and our constructed measure of municipality's exposure to the policy. Assuming the validity of the linear relationship, the estimates indicate that municipalities with exposure levels above 0.52 (which lies above the 95th percentile, as illustrated in Figure 4.1, Panel B) incurred negative monetary effects from the policy.[118]

The following models assume no parametric relationship between the amount reimbursed and the level of policy exposure faced by municipalities. Model 4 considers a very flexible specification by replacing the continuous measure of municipalities' policy exposure with quartile indicators of the distribution of this measure. Although municipalities facing significant constraints from the policy are shown to have experienced more limited increases in the amount reimbursed, the estimates indicate that they did not incur post-policy revenue losses. All municipalities, including those in the 4th quartile, saw, on average, increases in childbirth-related reimbursement following the policy implementation.[119] Finally, model 4 compares all municipalities where the threshold was binding to the those where it was non-binding and shows that the average post-policy change experienced by the former was, on average, larger for municipalities facing binding constraints.

In summary, the observed decrease in infant hospitalizations in municipalities with high exposure to the threshold policy occurred despite these areas having experienced lower

---

[117] This is computed by dividing the post-policy coefficient estimate (357,589) by the mean baseline outcome (703,420).

[118] This is recovered by solving the following identity 452,436 - 862,358*$EXP = 0$

[119] While increases were lower among municipalities facing levels of exposure above the median, as indicated by the negative magnitude of the interaction term coefficients, the overall changes were still positive for municipalities in the top quartiles. Estimates indicate that municipalities in the 4th quartile experience an increase in the reimbursed amount of approximately 280,480 - 276,589 = 3,891 relative to the baseline period.

increases in reimbursement revenues compared to those where the introduced threshold was less binding. Nevertheless, it's noteworthy that, overall, these municipalities did not experience revenue losses post-reform compared to the baseline amounts due to the concurrent upward adjustments in the unit tariffs of childbirth procedures, as described in Section 4.2.2. It is important to consider this background when interpreting our findings on health outcomes. Despite facing stricter constraints from the fixed reimbursement cap on the relative use of C-section deliveries, municipalities with a higher baseline propensity to perform such procedures had their financial stability largely safeguarded by these simultaneous governmental actions. The health consequences of the reform could have been different if hospitals had become financially strained because of the reimbursement caps.

Table 4.10: Policy effects on federal reimbursement of SUS deliveries

	Amount reimbursed for deliveries in SUS			
	(1)	(2)	(3)	(4)
		By Exposure to Policy		
	Post-policy avg. changes	Linear in exposure	Any exposure	By exposure quartile
$EXP * Pos$		-862,358***		
		(0.0000)		
$1_{EXP^{+}} * Pos$			168,238**	
			(0.0185)	
$1_{EXP^{+}\ in\ Q1} * Pos$				878,623***
				(0.0000)
$1_{EXP^{+}\ in\ Q2} * Pos$				198,309
				(0.2943)
$1_{EXP^{+}\ in\ Q3} * Pos$				-129,641***
				(0.0000)
$1_{EXP^{+}\ in\ Q4} * Pos$				-276,589***
				(0.0000)
Pos	357,589***	452,436***	280,480***	280,480***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N. obs.	5,520	5,520	5,520	5,520
Municipality FE	X	X	X	X
Baseline mean	*703,420*	*703,420*	*703,420*	*703,420*

Notes: Observations from SUS hospital claims are aggregated at the municipality-time level, where time refers to indicators of childbirth delivery during the 12-month periods before or after the policy announcement on May 29, 1998. The outcome variable concerns the amount reimbursed by the federal government for childbirth deliveries within the given municipality, measured in December 2022 prices. Model (1) simply regresses the outcome variable on the post-policy indicator. Model (2) is based on the estimation of Equation (4.3). Model (3) replaces the continuous values of the policy exposure measure with an indicator variable of positive values of exposure (i.e., EXP>0). Model (4) considers, instead, indicator variables for each quartile of the distribution of the exposure measure among municipalities where the threshold was binding. The corresponding model is described by Equation (4.4). The omitted category of all estimations refers to municipalities where the introduced threshold was deemed non-binding (i.e., EXP = 0). Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

4.8 Discussion

In a global context with limited evidence on effective strategies to address the escalating rate of C-sections, this chapter evaluates a distinctive policy intervention in Brazil - one of the countries with the highest C-section rates worldwide (Betran et al., 2021). In the late 1990s, the Brazilian government introduced a fixed monthly reimbursement cap on the percentage of C-sections among all births delivered in hospitals within the public health system. This cap was implemented alongside substantial increases in the reimbursed unit tariff of all childbirth procedures eligible for compensation, thereby alleviating potential financial burden on hospitals.

In a differences-in-differences approach, we exploit variation in the binding nature of the implemented cap, driven by largely diverse baseline propensities for cesarean procedures across municipalities. The key necessary assumption for the causal interpretation of our results is that municipalities facing tighter constraints due to the introduced threshold should not have been affected by external factors influencing decisions about childbirth procedures. This assumption would be violated if, for instance, childbearing mothers selected into less constrained municipalities in order to increase the likelihood of undergoing C-sections. Reassuringly, our robustness checks and event study analysis provide evidence against such hypotheses.

The results show that C-section use decreased more dramatically among municipalities where the policy was more binding. Heterogeneity analyses highlight pronounced effects among deliveries less likely to require medically justified C-sections, such as first births of younger mothers delivered from single pregnancies. Additionally, the decline in C-section likelihood was accompanied by a small decrease in the likelihood of low birthweight, suggesting that some C-sections eliminated by the policy might have been performed earlier than necessary.

Policy impacts on government-funded hospitalizations were also assessed. While we observe no effects on admissions of mothers due to post-partum complications, findings point to significant decreases in hospital admissions of children in their first year of life. These decreases were driven by respiratory-related causes, particularly chronic pulmonary disorders.

Results from the event study analysis and robustness checks provide evidence against the hypothesis that mothers shift to municipalities less constrained by the threshold when delivering their children and subsequently hospitalizing them in case of illnesses.

Additionally, evidence from placebo hospitalizations supports the understanding that mothers did not alter their likelihood of admitting their children to private hospitals as a response to the policy.

Importantly, our findings of health improvements among children less likely to be born via C-section align with a significant body of observational literature that establishes a link between potentially unnecessary C-sections and respiratory health deterioration in infants (Davidson et al., 2010; Håkansson & Källén, 2003; Hansen et al., 2008; Kristensen & Henriksen, 2016; Moore et al., 2012; Roduit et al., 2009; Salam et al., 2006; Thavagnanam et al., 2008; Tollånes et al., 2008), a relationship further validated by recent causal investigations (Card et al., 2023; Costa-Ramón et al., 2021; Jachetta, 2016). Finally, our results, showing no consequences for Apgar scores, align with mixed evidence found in other studies (Costa-Ramón et al., 2018; Currie & MacLeod, 2008).

Despite global efforts to reduce the rising prevalence of C-sections, most initiatives have shown limited effectiveness. One notable exception is a national reform implemented in Iran 16 years after the policy examined in this chapter. Pilvar and Yousefi (2021) reported a decrease of approximately 10% in the baseline national rate as a result of this reform. Interestingly, their result closely corresponds to the impact attributed in this study to a one standard deviation increase in our constructed measure of exposure to the evaluated threshold policy.

Among the policies examined thus far, the Brazilian reform assessed by this chapter stands out as the only one shown to protect mothers' health while improving their children's outcomes. While most studies examining health impacts found no effects on complications during delivery (Barili et al., 2021) and overall infant health (C. Melo & Menezes-Filho, 2023; Pilvar & Yousefi, 2021)[120], concerns have been raised about policies that strictly disincentivize C-sections inadvertently affecting medically beneficial cesarean procedures (Alexander, 2015; Barili et al., 2021; Berta et al., 2020).

The design of the policy is likely to have contributed to its success. While imposing stringent constraints on the relative use of C-sections among providers with a high propensity to perform them, coordinated actions were taken to safeguard the financial stability of hospitals. Crucially, generous increases in the unit fee of reimbursable C-sections

[120] Although C. Melo & Menezes-Filho (2023) find a slight increase in the 1st minute Apgar score, such increase disappears by the 5th minute after birth.

(approximately 50%) were implemented to protect the revenue of municipalities more constrained by the reimbursement cap. Since the compensation increase for vaginal deliveries was even higher (around 70%), providers with baseline C-section rates below the cap did not face incentives to increase the adoption of the surgical alternative. The implementation of the policy resulted in an additional cost of R$ 240 million (as of May 1999) for the federal government during the 12 months following its implementation, representing a 55% increase compared to the preceding 12 months.[121]

Moreover, the changes in financial incentives introduced by the policy did not directly impact the remuneration of physicians, who ultimately make procedure decisions. Although physicians may respond to the incentives of the institutions employing them, they might do so to a lesser extent if paid fixed salaries. Since they do not directly bear financial consequences for their choices, physicians are more likely to act against incentives to safeguard C-sections they deem beneficial for their patients. Lastly, the cap was set at a reasonably high level, exceeding the baseline average rate of targeted hospitals. Consequently, the facilities subjected to the binding cap were those with excessively high C-section rates that were generally difficult to justify on medical grounds. Indeed, our finding that declines in infant hospitalization were concentrated among municipalities in the top quartile of our policy exposure measure suggests that C-sections detrimental to patient health were systematically performed in municipalities with a high baseline propensity for such procedures.

A primary limitation of our study arises from data constraints, preventing an assessment of the policy's impact among the targeted hospitals. As previously highlighted, post-threshold implementation, the reliability of SUS hospital claims is compromised due to the lack of data on non-reimbursable C-sections and potential manipulation by providers aiming to maximize reimbursement. To overcome this limitation, our analysis relies on birth certificates. They consist of the universe of all births in the country, but do not include identifiers of hospitals of birth. While this may be seen as a constraint, it yields robust and conservative results. Results accommodate within-municipality hospital selection (i.e., shifts between hospitals located in the same municipality) in response to the policy. Additionally,

[121] This is equivalent to ~$50 million in prices of 1999 (based on exchange rate of 1 BRL = 0.204271 USD in this period) or ~$87 million in prices of 2022 (based on cumulative inflation of 75.7%).

because our analysis includes private sector births not directly targeted by the policy, estimates should be interpreted as conservative lower bounds.

The key message for policymakers in this chapter is that, in countries grappling with overuse of C-sections, implementing stringent supply-side constraints on their utilization could offer a viable solution. While policies directed at the demand side have proven highly ineffective (De Oliveira et al., 2022; C. Melo & Menezes-Filho, 2023; L. Melo, 2021), efforts to shift supply-side incentives away from C-sections have shown limited effectiveness (Barili et al., 2021; Berta et al., 2020; Keeler & Fok, 1996; Kozhimannil et al., 2018; Lo, 2008). While these studies have shown that lowering the relative reimbursed fees for C-sections compared to vaginal deliveries shows little impact, the outright refusal of reimbursement, as indicated by this study, effectively reduces the use of cesarean procedures. Nevertheless, safeguarding medically justified cesarean procedures should always remain a priority.

A criticism of the threshold implemented by the Brazilian reform is that the threshold was uniformly implemented across hospitals seeing patients with different risk factors and, therefore, medical need for C-section. A fairer and potentially more advantageous approach would have been to implement hospital-specific quotas based on their respective patient populations, rather than uniformly imposing a fixed quota nationwide. That said, our results reveal that the implemented threshold policy was not harmful, on average, to patient health. This could be because the threshold was set high enough to protect the maximum share of patients expected to need a C-section. Alternatively, physicians might have conducted C-sections whenever there was a medical indication to do so, regardless of the reimbursement cap in place, while restricting reductions of C-sections to cases when this type of delivery was not necessary. As argued before, the fact that their remuneration was not affected by the policy might have kept their incentives reasonably aligned with those of patients.

5 INEQUALITIES IN BIRTH TIMING MANIPULATION

The chapter aims to elucidate key factors influencing treatment timing decisions in both the public and private healthcare sectors. Distinct health systems can significantly influence not only the degree to which treatment timing is manipulated but also the underlying motives behind it, whether driven by medical need or pure convenience, resulting in potential equity consequences. Understanding these dynamics is crucial for assessing scope for policy intervention or regulatory measures. The analysis in this chapter focuses on childbirth delivery, for which we observe the universe of procedures performed in the country for over a decade.

Human birth is naturally programmed to occur via vaginal delivery under spontaneous labouring. However, medicalization in delivery has increasingly altered the nature's uniform distribution of births over time, across days or even within days, either because of appropriate manipulation of the timing of birth aimed at minimizing medical risks, or because of non-clinical reasons determined by convenience and opportunistic behaviour. While there is research documenting the role that medical appropriateness and convenience reasons play in the medicalization of delivery and in the manipulation of the timing of birth (Becker, 2007; Gijsen et al., 2012; Gould et al., 2003; Hong et al., 2006; Jensen & Lorch, 2017; Lo, 2003; Lyndon et al., 2015; Palmer et al., 2015; Zampieri et al., 2018), equity concerns remain largely unexplored.

Inequalities across different population groups may arise from both sources of manipulation. Patients with access to appropriate medicalization are more likely to benefit from manipulation aimed at risk minimization, designed to protect patient health. In particular, socioeconomically vulnerable parents may be relatively more constrained in the access to quality hospital services for reasons such as financial constraints, distance to available care and discrimination at admission (Okeke & Chari, 2018; Slaughter-Acey et al., 2019; Treacy et al., 2018).[122] Additionally, women from racial/ethnic minority groups and from low socioeconomic background tend to be less involved in the decision-making process during pregnancy and delivery (Altman et al., 2019; Attanasio et al., 2018), as well as are more likely to experience disrespect and abuse during childbirth (Leal et al., 2017; McLemore et

[122] Sosnaud (2021) shows that racial disparities in neonatal mortality in some US states persists after controlling for differences in birthweight distribution and socioeconomic characteristics. The author reads his findings as indicative of differential receipt of appropriate medical care during and after birth, either in access or quality.

al., 2018; Vedam et al., 2019).[123] Mothers facing higher bargaining power are more adept at influencing manipulation decisions driven by convenience, by asserting their preferences or safeguarding against self-interested physician choices.

While previous research has explored inequalities in medical treatment within the context of childbirth, studies have predominantly focused on access to institutional care (Okeke & Chari, 2018; Slaughter-Acey et al., 2019; Treacy et al., 2018) and procedure choices (Robinson et al., 2023; Valdes, 2021). Racial patterns in the timing of births have been largely overlooked. Additionally, there is limited understanding of the extent to which incentives embedded in health systems might mitigate or exacerbate inequalities during childbirth. In a recent study, Ferraro et al. (2021) argue that convenience significantly influences the choice of childbirth procedures in the Argentinian private sector but holds less sway in the public sector.[124]

Using data from approximately 37 million births in Brazil, this chapter investigates the manipulation of the timing of births around types of days associated with different incentives. Specifically, we investigate birth patterns in the vicinity of inauspicious dates (inconvenient to parents), of days on which the Brazilian Congress of Obstetricians and Gynaecologists is held (inconvenient to physicians), and of bank holidays (typically inconvenient to physicians, but also times when hospital resources might be scarcer, and risk is expectedly higher).[125] We explore birth timing manipulation separately for two population groups, black and white mothers, in both the private and public health systems.[126]

Brazil is a unique empirical setting as socioeconomic inequalities and health system segmentation allow variation across those different margins to manifest. This enables us to uncover whether patterns in inequalities between black and white mothers in birth timing exist, and how they behave within different hospital systems. Incentives within public and

---

123 Greenwood et al. (2020), for instance, document evidence that black newborns cared by black physicians experience improved chances of survival.

124 The authors report that, in private hospitals (but not in public ones), a higher likelihood of C-section occurs when women go into labour on business days compared to weekends and bank holidays. They attribute the observed pattern in the private sector to motivations rooted in convenience.

125 Although risk of delivering during congress days might also increase given that higher skilled doctors are usually more likely to unavailable, risk increase is expectedly lower than that associated with bank holidays. This is because congress events are unlikely to substantially affect the operation of hospitals. Additionally, women who would have had births delivered by high-skilled doctors who become unavailable because of the congress are likely to be directed to other skilled physicians not attending the conference.

126 Our classification of black mothers comprises all women of colour. Black/brown mothers constitute 98% of the sample, while the remaining 2% is composed of women who identify as yellow or indigenous.

private hospitals are remarkably different. Public hospitals have 100% of the beds affiliated to the publicly funded healthcare system (*Sistema Universal de Saúde*, SUS), which is universal in coverage.[127] Physicians are usually employed in SUS under contracts specifying a fixed number of hours and salaries per month. While similar contracts also exist in the private sector, it is very common for physicians to have multiple employment attachments based on reduced number of hours, sporadic shifts, or on-call services (Costa et al., 2022). As a result, many physicians in the private sector act as autonomous doctors with less formal employment relationships and tend to enjoy more freedom to decide on how to organize their schedule. The remuneration of physicians in the private sector varies between fixed salaries and fee-for-service schemes, depending on their specific employment attachments.[128]

Mothers who use the public sector do not incur in any cost. They cannot choose the physician who will assist them according to their own preferences.[129] This makes the patient-physician relationship less personal and centralizes the decision-making at the time of delivery in one agent, typically the physician in charge at the hospital. Mothers who opt for the private sector usually have their delivery fees covered by the private insurance (if insured). They are mostly free to choose the professionals who will follow their gestational period and deliver their babies.

In order to investigate how the number of births, mode of delivery, and risk profile of births are distributed in the vicinity of inconvenient days, we used a panel of data at the hospital-day level, over the 2006-2019 period, and relied on a fixed effects regression specification. While all inconvenient days have some fixed time parameter, they occur on different days of the week or month across different years, thus enabling us to recover causal effects conditional on time and hospital fixed effects. Results are presented separately for black and white mothers, public and private sectors. We also report the racial gap in birth timing manipulation by examining the difference in the number of births between black and white mothers for which timing is altered within hospitals exclusive to either the public or private sectors. Finally, we turn to hospitals serving both the public and the private sectors

[127] In the remainder of this chapter, we define *public* (*private*) hospitals as those with 100% (0%) of obstetric beds reserved to SUS.

[128] Hospitals serving SUS have procedures reimbursed by the federal government based on a fee-for-service model combined with fixed-budget payment scheme (Levin, 2006). In the private sector, fees vary according to insurance plans.

[129] They are usually assisted by physicians and nurses during the prenatal period, and typically have a different doctor for their delivery (Domingues et al., 2014).

during the observed time horizon to investigate changes in such racial disparities as the *same* hospital becomes more reliant on public funding.[130] This enables us to obtain effects for different levels of public sector exposure that are not driven by hospital heterogeneity.

We present evidence of birth timing manipulation in both sectors, with a notably higher prevalence among white women delivering in private hospitals. Our findings suggest that convenience and larger bargaining power of white women drive manipulation in the private sector, whereas at least part of the manipulation in public hospitals reflects medical appropriateness and results in reductions in racial disparities in quality of received care. At times of expectedly lower quality of service delivery, manipulation in the public sector is especially targeted at births from black mothers and riskier pregnancies. A similar pattern is observed within the same hospital, as their funding becomes more attached to SUS over the months of our time sample.

The results from our analysis are relevant for the optimal design of incentives in health systems and the regulation of medical practices. More generally, the extent to which the manipulation of birth timing is determined by individuals' choice and health system characteristics may open the possibility of inequalities not only in birth timing itself, as examined in this study, but also in the effects that manipulation unrelated to medical reasons might have on newborn health and its associated long-term consequences.

5.1 Data

The main source of data is the Brazilian National System of Information on Birth Records (Datasus/SINASC), which includes the universe of all registered live births in Brazil at the birth level. In addition to individual-level characteristics, SINASC contains information on date of birth, hospital code, and mode of delivery. We also used information from SINASC to classify high-risk deliveries as those involving multiple pregnancies, newborn with congenital anomaly, or newborn in breech or shoulder positions before birth.

We identified whether the hospital is affiliated to the public healthcare system or not by matching birth-level records with monthly data from the National Registration of Health Facilities (Datasus/CNES). We used the hospital code and month of birth as key variables

[130] This study uses terms such as racial gradient, racial gap, racial differences, racial disparities, and racial inequalities interchangeably when describing differences in patterns in which the timing of deliveries from black mothers, as compared to white mothers, is manipulated around inconvenient days within the same hospital-day.

for the linkage. CNES contains monthly information on the number of obstetric beds by type of funding for every registered hospital in the country. This allowed us to compute the share of obstetric beds affiliated to SUS in each hospital at every month. Our analysis encompasses the period between 2006 (the year when Datasus/CNES was made available) and 2019 (the year prior to the Covid-19 pandemic).

The original SINASC dataset contains the universe of 40,037,255 registered hospital births during the period from 2006 to 2019. We keep observations for which the hospital identifier is informed (and found in the Datasus/CNES database) as well as the race of the mother. We end up with 37,069,691 births taking place in 5,752 hospitals. In most of our analyses, we restrict our sample to births occurring in hospitals which were either exclusively public or exclusively private. We classify public (private) hospitals as those with all (none) of the available obstetric beds attached to SUS during the entire period of analysis.[131] In this restricted sample, we observe 16,072,998 births which took place in 2,900 public hospitals and 5,656,002 births which happened in 875 private hospitals.[132]

Table 5.1 presents summary statistics for all observed births as well as for the sample of births occurring in exclusively public *vs* private hospitals. We observe more births taking place in public hospitals, which is consistent with the fact that the number of public hospitals is approximately three-fold the number of private ones. The average number of daily deliveries per hospital indicates a concentration of black women delivering in public hospitals and white women in private facilities. Whereas white women in our sample split equally between public and private facilities, the large majority of black mothers (87%) deliver in public hospitals. The share of C-sections is higher for white mothers in both types of hospitals, and generally prevalent in private facilities, representing more than 85% of all births. This is more than double the share of C-sections among public hospitals, which is of approximately 40%. We observe a slightly higher share of riskier deliveries in private hospitals, both among white and black mothers.

131 In other words, for these analyses, we dropped births that occurred in hospitals that served simultaneously the public and the private sectors for at least one month during the entire period.

132 When collapsing the data by hospital-day, we input zero births per day for all days with no reported births during the period between the first and last reported births.

Table 5.1: Summary statistics by hospital type and race of the mother

	All	Public	Private	Diff. (Pub-Pri)
N. of Hospitals	5,752	2,900	875	-
N. of Births	37,069,691	16,072,998	5,656,002	-
Panel A: Black				
N. of Births	21,886,549	12,122,752	1,750,181	-
Avg. N. births/hospital-day	0.93 (2.39)	1.02 (2.53)	0.58 (1.48)	***
% C-section	0.46	0.38	0.85	***
% High-risk births	0.05	0.05	0.06	***
Panel B: White				
N. of Births	15,183,142	3,950,246	3,905,821	-
Avg. N. births/hospital-day	0.65 (1.76)	0.33 (1.16)	1.30 (2.96)	***
% C-section	0.64	0.43	0.87	***
% High-risk births	0.05	0.05	0.06	***

Notes: The second column (i.e., All) refers to all births for which we observe hospital identifier and race of the mother. The following columns present the samples of births in hospitals which are exclusively public or exclusively private, defined as those with, respectively, 100% and 0% share of obstetric beds affiliated to SUS, the Brazilian public healthcare system, throughout the period between the years of 2006 and 2019. Panels A and B, highlighted in grey, break the sample by race of mother. The table reports the total number of hospitals, total number of deliveries, mean and standard deviation (in parenthesis) of number of deliveries by hospital-day, as well as share of births delivered by C-section, and share of high-risk deliveries (as those involving multiple pregnancy, non-cephalic foetus presentation or newborn with congenital anomaly). The last column indicates whether the differences of mean values between public and private hospitals are statistically significantly different from zero, where *** corresponds to p<0.01.

5.2 Methods

In order to investigate how the number of births is distributed in the vicinity of inconvenient days, we used a panel of data at the hospital-day level, over the 2006-2019 period, and relied on a fixed-effects regression specification. Our model follows below:

$$Y_{hymd} = \alpha + \sum_{i=1}^{I} \beta_i (d_i)_{ymd} + \sum_{\substack{n=-9 \\ n\neq 0}}^{+9} \gamma_n (d_n)_{ymd} + \rho_{ymd} + \theta_m + \eta_y + \phi_h + \varepsilon_{hymd} \quad \textbf{(5.1)}$$

where *h* indexes hospital, *y* indexes year, *m* is a subscript for calendar month, and *d* is a subscript for day. We run the model separately for each class of inconvenient day (inauspicious, congress, one-day bank holiday, and two-day bank holiday), type of hospital (public *vs.* private), and race of the mother (white *vs.* black). The term d_i refers to a dummy indicator variable representing inconvenient day. While inauspicious days and one-day bank holidays constitute single-day periods, congress and two-day bank holidays involve consecutive days in a row. In the equation above, we specified $I = 1$ for one-day bank holidays and inauspicious days, $I = 2$ for two-day bank holidays, and $I = 4$ for congress days.[133] We estimated coefficients for dummies indicating each day around these inconvenient periods within a time window spanning from 9 days before the inconvenient period and 9 days after the inconvenient period, thus covering approximately three weeks. The term $d_{(n)}$ refers to a dummy that indicates the n[th] day before the beginning of the inconvenient period (for negative values of *n*) or after the inconvenient period ends (for positive values of *n*). The remaining terms ρ_{ymd}, θ_m, η_y, and ϕ_h refer, respectively, to day of the week, month, year, and hospital of birth fixed-effects. The exact day of the week or month on which a given inconvenient day falls each year varies across years, thus allowing us to use a range of time fixed-effects to absorb any confounding seasonal influence on the evaluated outcome.[134] Hospital fixed-effects adjust for persistent determinants of demand for hospital services and of quality of care at the hospital level. Finally, the error term is represented by ε_{hymd}. We cluster standard errors at the hospital level to account for serial correlation within hospitals over time.

These results will enable an exploration of how birth patterns evolve during the time surrounding inconvenient days compared to the expected patterns in the absence of such days, based on within-hospital activity, conditional on year, calendar month, and day of the week. The rationale behind evaluating patterns over a relatively wide time window is to enable a comprehensive analysis of the broader repercussions of inconvenient dates. The shifting of delivery timings may extend beyond those that would have happened during these exact dates. Due to increased hospital activity in the surrounding proximity of inconvenient dates, the timing of other births could also be affected. Moreover, the use of wide time frames should be seen as an overall conservative approach. This is because the days around the inconvenient days explicitly modelled in the regression are, by design, excluded from the reference categories in the regression analysis. If activity during these days is indeed influenced by the inconvenient dates, excluding them from our time window could introduce

---

133 For the year of 2007, the Brazilian congress lasted for 5 (instead of 4) days. Our estimate on the 5th congress day will, therefore, be very imprecise.

134 Most inauspicious dates have fixed month and day of month each year (April 1, February 29) but occur in different days of the week across different years. The same happens with most bank holidays (e.g., Independence Day). Other inconvenient days have fixed weekdays but may fall in different months across years (e.g., Carnival, Friday the 13th). Congress days have fixed month (November, every two years) but different weekdays across different years.

bias into our analysis. Conversely, their inclusion provides valuable information on whether or not they are affected.[135]

In Equation (5.1) above, our main outcome variable Y_{hymd} refers to the number of births at the hospital-day level, which was computed for different sub-samples (private *vs.* public, white *vs.* black mothers). We also specified analogous models at the birth-level for the likelihood of C-section and for the risk indicator in order to characterize changes in the mode of delivery and risk profile around inconvenient periods. In both cases, the outcome variable is binary, and we relied on linear probability models.

Finally, we study racial gradient in manipulation away from inconvenient days. We begin by estimating Equation (5.1), at the hospital-day level, for the difference in number of births between black and white mothers. To examine whether racial gap manipulation differs across sectors, we estimate separate regressions for public and private hospitals. Next, we turn to the universe of all births and exploit variation from hospitals having experienced different degrees of affiliation to SUS over the time horizon of our analysis. Because the monthly share of obstetric beds varies over time for these hospitals, we investigate whether manipulation dynamics between black relative to white mothers change over time as the same hospital becomes more (or less) attached to the public sector.

Based on the hospital's monthly share of obstetric beds associated to SUS, we create indicator variables of intensity of exposure to SUS representing whether hospitals were exclusively private (0% SUS beds), predominantly private (share of SUS beds above 0 and below ¼), mostly private (share of SUS beds above or equal to ¼ and below ½), mostly public (share of SUS beds above or equal to ½ and below ¾), predominantly public (share of SUS beds above or equal to ¾ and below 1), or exclusively public (100% SUS beds). In the model below, these indicator variables are represented by $D(SUS_r)$, where r indicates each one of these categories.[136] These indicator variables are then interacted with dummy indicator variables of inconvenient periods and nearby days (9-day period before and after the inconvenient period).

[135] Note that Jacobson et al. (2021) considered an even larger time window of 28 days when assessing birth timing patterns around US holidays.

[136] The indicator variable $D(SUS_r)$ is equal to1 if the share of obstetric beds in hospital *h* affiliated to SUS during year/month *y*/*m* lies in category *r*, and 0 otherwise.

$$
\begin{aligned}
Y_{hymd} = \alpha &+ \kappa D(Incon)_{ymd} + \gamma D(Around)_{ymd} + \sum_r \tau_r D(SUS_r)_{hym} \\
&+ \sum_r \beta_r \{D(SUS_r)_{hym} * D(Incon)_{ymd}\} \\
&+ \sum_r \delta_r \{D(SUS_r)_{hym} * D(Around)_{ymd}\} + \rho_{ymd} + \theta_m + \eta_y \\
&+ \phi_h + \varepsilon_{hymd}
\end{aligned}
\quad \textbf{(5.2)}
$$

Equation (5.2), above, describes the second regression model we estimate. Another difference from the previous model is that it aggregates all days falling during inconvenient periods or nearby periods in single indicators. The indicator variable $D(Incon)_{ymd}$ is a dummy variable for inconvenient days, while $D(Around)_{ymd}$ is an indicator for the days preceding the start of the inconvenient period by up to 9 days or posterior to the end of inconvenient period by up to 9 days. This alleviates the more limited variation available once the interaction terms are introduced to the model. Besides, we are mostly interested in understanding overall manipulation away from inconvenient periods (instead of the dynamics of manipulation around inconvenient periods). Our coefficients of interests, β_i's, inform the heterogeneity in manipulation between black and white mothers when the given hospital becomes more or less attached to the public health system.

We consider three types of inconvenient days: inauspicious days, dates during the Brazilian Congress of Obstetricians and Gynaecologists, and bank holidays. Inauspicious days are classified as those that fall on Friday the 13th, April Fools' Day (April 1) and Leap Day (February 29). While Friday the 13th is well-known as a day of bad luck, kids born on April Fools' Day are more likely to be stigmatized and those born on Leap Day can only celebrate their birthdays on leap years. Given the stigma and inconvenience attached to these dates, mothers have incentives to avoid delivering on such days.[137] The Brazilian Obstetrician-Gynaecologist Congress generally happens every two years and lasts for approximately four days.[138] Doctors who decide to attend the conference would not be available to deliver births during the days when the event takes place. This may lead to

[137] Bullying among schoolchildren is very common in Brazil (Mello et al., 2017) as well as it is the stigma attached to these days. For instance, see https://www.enfoquems.com.br/como-surgiu-a-crenca-de-que-a-sexta-feira-13-e-um-dia-de-azar/ (URL link in Portuguese)

[138] The location where the congress takes place varies across years and usually corresponds to one of the 27 state capitals in the country.

manipulation driven by the doctor, and an increase in the risk of delivering in such dates given restricted supply and selection of available providers.[139] Finally, bank holidays may trigger manipulation by both parents and providers due to either convenience or risk aversion given that quality of hospital service delivery during national holidays tend to be poorer (i.e., higher procedural risk due to restricted supply and lower skills of available professionals, longer waiting times, etc).

During the time span considered in our analysis, there have been 7 congress years comprising of 29 congress days along with 43 inauspicious days and 137 bank holidays. As regards to bank holidays, we considered solely those falling between Monday and Friday. We identified 107 one-day and 15 two-day holidays (mostly during Carnival). The list of all dates can be found in Table D.1, in the appendix.

5.3 Results

5.3.1 Birth timing manipulation

We present our main results of the dynamics around inconvenient days in coefficient plots. Point estimates and their respective 95% confidence intervals are presented in plots separately for births that occurred in public (left-hand plots) and in private facilities (right-hand plots). Each plot documents the results of two different regressions: estimates for the sample of white women are presented in grey while those for black woman appear in black.[140] For ease of visualization, when presenting results, we refer to coefficients preceding and succeeding the inconvenient days as $\pm d_{|n|}$ in lieu of d_n, as referred to in Equation (5.1).[141] Plot scales are fixed for each outcome variable. Outcome averages are informed in Table 5.1.

Before presenting our results, it is important to acknowledge that shifts in the timing of births away from inconvenient days may be accompanied by corresponding changes in the method of delivery in either direction or have no impact at all. Births that would occur via vaginal delivery during inconvenient days may be delivered earlier via C-sections. Conversely, those originally planned as C-sections could potentially occur vaginally following

[139] There is less scope for manipulation away from conference days to be driven by parents due to risk aversion given that these conferences are not advertised to the public and are unlikely to be disseminated by word of mouth given its targeted audience and infrequent nature (4/5 days every 2 years).

[140] We report all regression results (coefficients and standard errors) in the appendix (Table D.2-Table D.5).

[141] We replace d_n by $+d_n$ for positive values of *n* and by $-d_{|n|}$ for negative values of *n*.

labour onset after the inconvenient date. Alternatively, C-sections might be simply rescheduled from inconvenient dates to other nearby days.

Figure 5.1 reports the birth timing dynamics around inauspicious days. In Panel A, we observe that the total number of deliveries remains unaltered in public hospitals while there is a substantial reduction in births in private facilities, particularly among white mothers (d_1 = -0.269, 95% CI -0.3145 to -0.222). The point estimate of the exact inauspicious day corresponds to 21% of the daily hospital average for the sample of white mothers and 11% for the sample of black mothers who deliver in private hospitals.[142] The average number of daily births of white (black) women in private hospitals is 1.3 (0.58), as informed in Table 5.1. The remaining coefficients indicate that deliveries are displaced to days in the vicinity of the inauspicious period. These results are consistent with the fact that private settings provide women with the possibility to choose the physicians that align with their preferences.[143] Besides, our finding that birth timing manipulation is greater for white women (in absolute and relative terms) in a context where it is on the interest of mothers to move the day of their deliveries reinforces the understanding that skin colour is an important factor in determining the extent to which their voices influence the decision-making process during prenatal care and childbirth.[144] Panel B suggests that at least part of such manipulation is made possible with the use of C-sections. We observe a lower likelihood of that type of delivery on inconvenient days, but a slightly increased likelihood on vicinity days, indicating that births that would have been delivered surgically on the inauspicious dates are rescheduled to the period before or happen after (either through a postponed C-section or after spontaneous labour). As mentioned above, rises in the number of births following the inconvenient period may not necessarily be accompanied by a higher likelihood of C-sections if it occurs spontaneously after the onset of labour. In Panel C we do not observe any systematic changes in the pattern of riskier deliveries, which suggests that manipulation is unrelated to the risk profile of births.

---

[142] The coefficient of inauspicious day for the sample of black mothers who deliver in the private sector is d_1 = -0.066 (95% CI -0.081 to -0.051) and is reported in Figure 5.1 (in black) and Table D.2 (Panel A, last column).

[143] Furthermore, women in the private sector have more time and opportunity to influence the decision on the day of birth as the obstetrician who assists their delivery is usually the same professional that has treated them during prenatal care. They are also more likely to deliver through C-section (see Table 5.1), thus having the possibility to schedule their childbirth from the start.

[144] Provided that preferences of white mothers are similar to those of black mothers.

We then assess physician-driven manipulation by examining the period around the Brazilian Obstetrician-Gynaecologist Congress. Panel A in Figure 5.2 shows a substantial drop in the number of deliveries during congress days in private hospitals. This is expected given that women who deliver privately tend to give birth with the same obstetrician who had cared for them throughout the prenatal period (Domingues et al., 2014), while in the public sector delivery is typically carried out by the professional in charge. Therefore, time manipulation likely accommodates availability restrictions of the chosen physician in the private sector. We observe that some manipulation does occur in deliveries of white and black mothers in the public sector, but without significant disparities between groups. Yet, differences are relatively well-marked among white mothers in private facilities and might reveal a tighter coordination of preferences between white mothers and physicians.[145] Expectant mothers may prefer to have their babies delivered by the physician who accompanied them throughout pregnancy a few days before the original plan over seeing another provider with whom they have no established relationship. Physicians, who can only schedule a limited number of deliveries on the days before the conference, are likely to give priority to white mothers, who supposedly enjoy greater bargaining power.[146] Another possibility is that obstetricians typically chosen by white mothers are more likely to attend congress events than those who deliver the babies of black mothers.[147] Both hypotheses would suggest racial disparities, either in terms of priority given by physicians to white mothers in the rescheduling of deliveries (i.e., within-doctor selection) or in terms of selective matching between white mothers and higher-skilled doctors (i.e., physicians who attend conferences tend to be more informed and better trained). Because C-section is widespread in the private sector (and slightly more prevalent among white women), time manipulation of births with due date around congress days would not necessarily entail a change in mode of delivery. Indeed, we observe a slight decrease in the likelihood of C-sections during congress days, with a corresponding increase mainly before that period. This

[145] The coefficient around the middle of the time window of conferences is d_3 = -0.275 (95% CI -0.326 to -0.224) for white women in the private sector, whereas it is d_3= -0.109 (95% CI -0.136 to -0.082) for black mothers. In public hospitals the respective estimates are d_3=-0.013 (95% CI -0.023 to -0.003) and d_3=-0.026 (95% CI -0.044 to -0.009). The average numbers of daily births by race of women and type of hospital funding in our sample are informed in Table 5.1.

[146] When expectant mothers choose their obstetrician, around the 4th month of pregnancy, they are unlikely to be aware of daily time constraints of physicians many months ahead (and, probably, so are the physicians themselves). Besides, it is hard to precisely predict the due date of delivery at the time of the initial consultation with the obstetrician.

[147] White mothers may be less constrained in their choice of obstetrician if they have more flexible insurance plans.

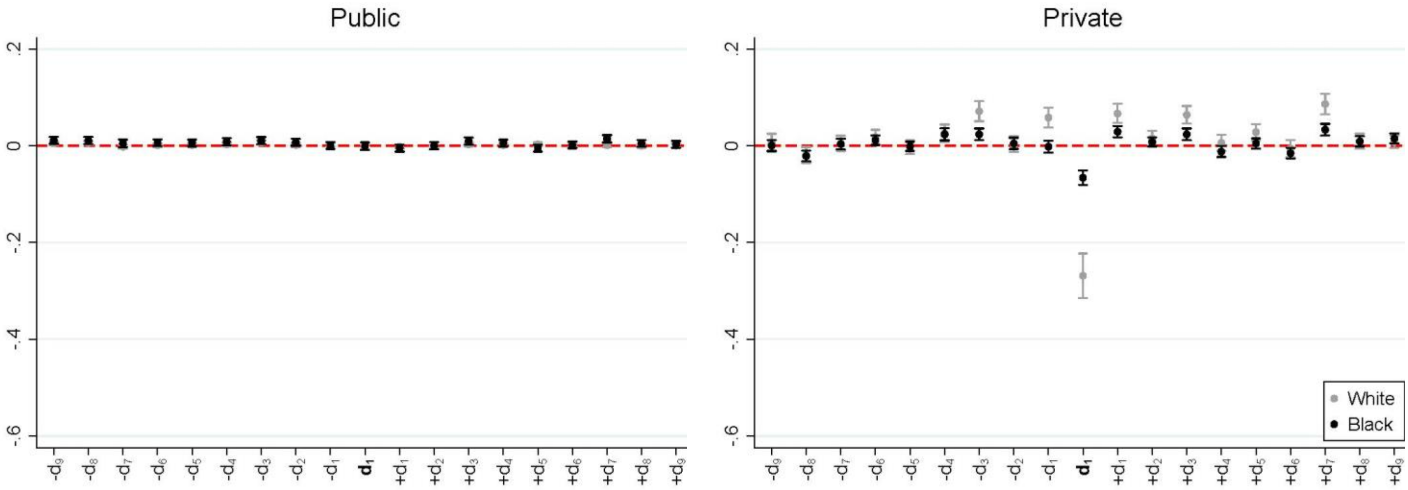
suggests that C-sections took place a few days before they would, had it not been for the conference itself. Finally, once again there is not any clear patterns related to the risk profile of births.

Figure 5.3 and Figure 5.4 report the results on birth timing around one- and two-day bank holidays[148], which are in principle inconvenient both for parents and physicians, and also potentially riskier as facilities may be understaffed. In Panel A of Figure 5.3, we observe again a reduction in the number of deliveries during holidays both in public and private hospitals, but substantially more salient among white mothers in the private sector (d_1= -0.622, 95% CI -0.706 to -0.538) -- point estimates are roughly two-fold those reported in Figure 5.1 and Figure 5.2. Panel B documents a significant decrease in the likelihood of C-sections during the inconvenient period, accompanied by an increased likelihood on vicinity days, especially before holidays. In Panel C we now observe some systematic variation in the risk profile patterns. While the dynamics remain unaltered in the private sector (white mothers: d_1=+0.000, 95% CI -0.002 to +0.003; black mothers: d_1= +0.000, 95% CI -0.003 to +0.004), there is a decrease in the share of riskier deliveries during bank holidays in the public sector for both white and black women (white mothers: d_1= -0.005, 95% CI -0.006 to -0.003; black mothers: d_1= -0.004, 95% CI -0.005 to -0.002). Given the average share of high-risk births in public hospitals, the coefficients point to a decrease of 8-10% in the likelihood of high-risk deliveries taking place during bank holidays. We observe qualitatively similar patterns related to two-day holidays (see Figure 5.4).

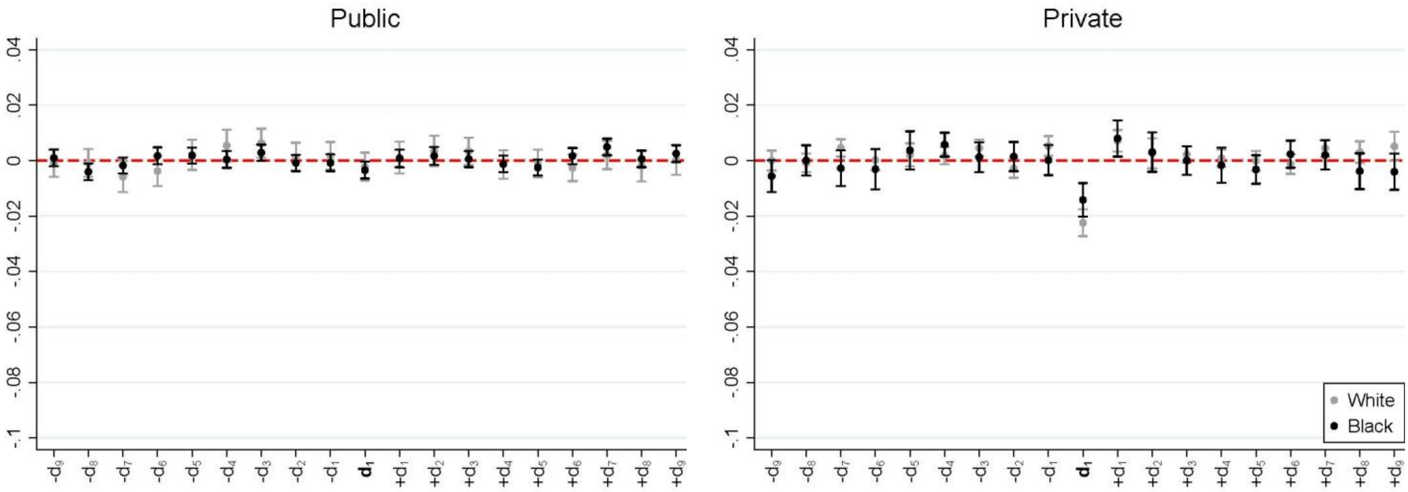
[148] We present results separately for one- and two-day bank holidays as the dynamics may change according to the duration of bank holidays.

**Figure 5.1: Regression coefficients of days around *inauspicious dates*:
Number of births, delivery type, and risk profile**

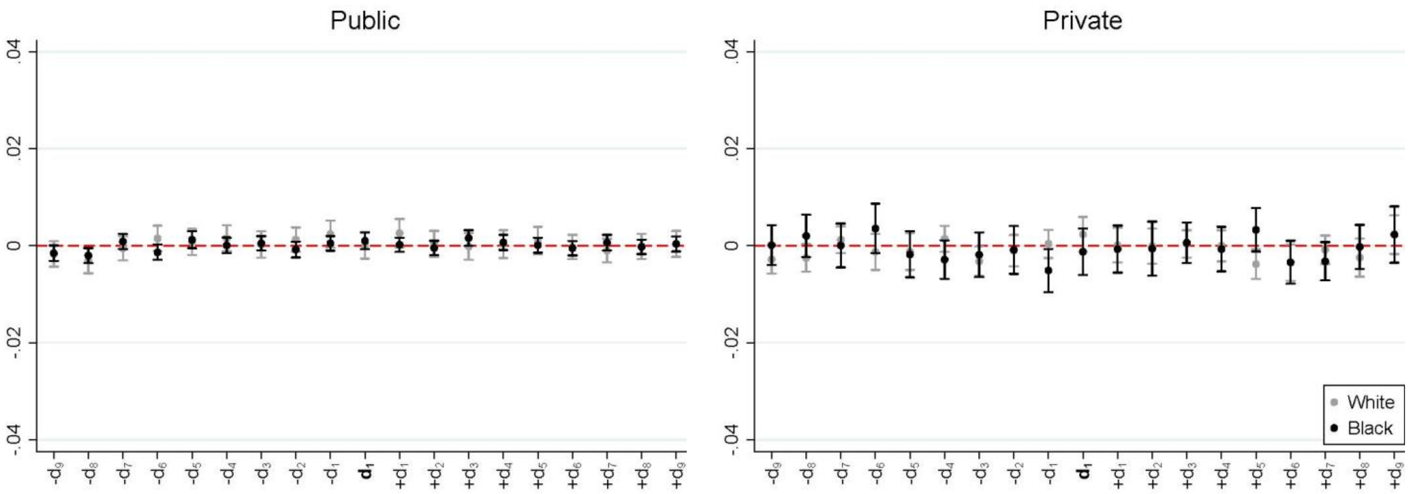
Panel A: Number of deliveries



Panel B: Likelihood of C-section



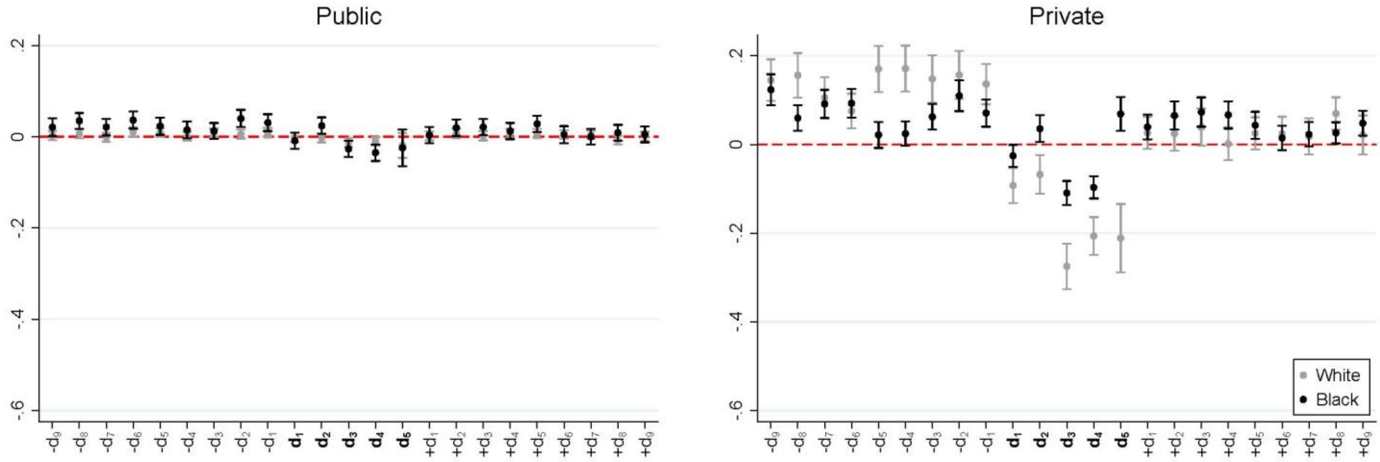
Panel C: Risk indicator



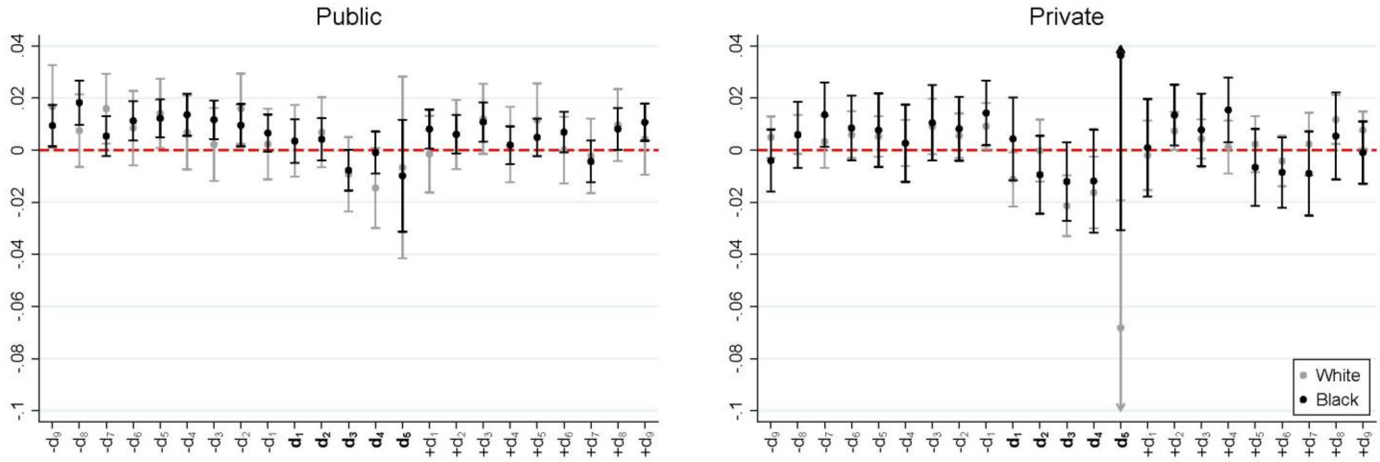
Notes: Each plot shows point estimates and respective 95% confidence intervals for two separate regressions: one for the sample of white women (in grey) and another for the sample of black woman (in black). Panel titles indicate outcome variables. Unit level is individual birth for regression results reported in Panels B and C and hospital-day-race for those reported in Panel A. Risk indicator is classified as multiple pregnancy, non-cephalic foetus presentation or newborn with congenital anomaly. Standard errors are clustered at the hospital level. For more details on regression specification, see Equation (5.1) in Section 5.2. The coefficient term -d_n (+d_n) refers to the n^th day before (after) the *inauspicious day*, represented by d_1. Coefficients' point estimates and respective standard errors as well as the number of observations of each regression can be found in Table D.2.

Figure 5.2: Regression coefficients of days around *Obstetricians-Gynaecologists Congress*: Number of births, delivery type, and risk profile

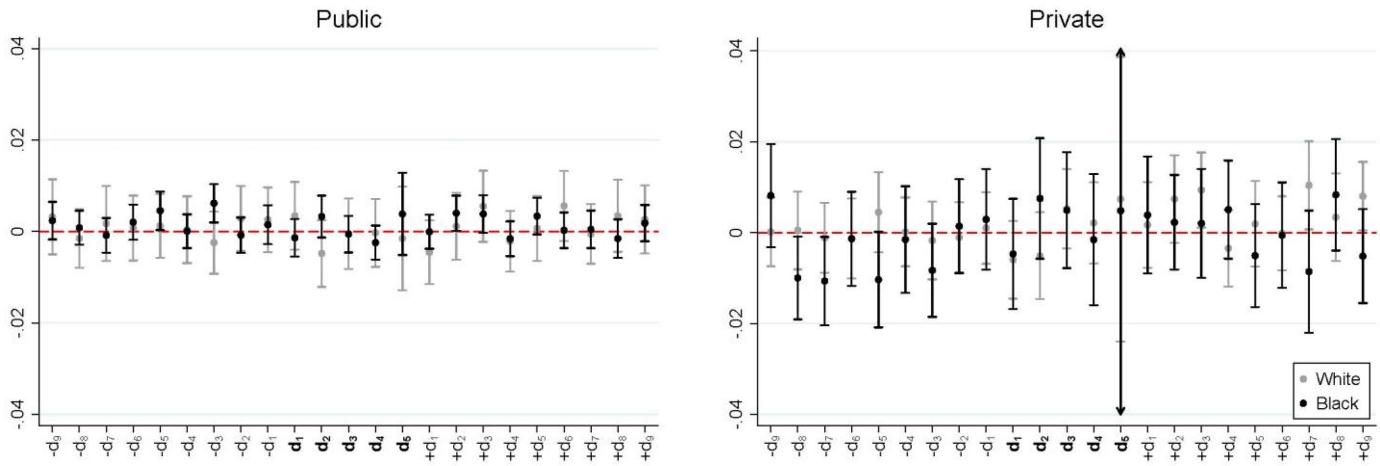
Panel A: Number of deliveries



Panel B: Likelihood of C-section



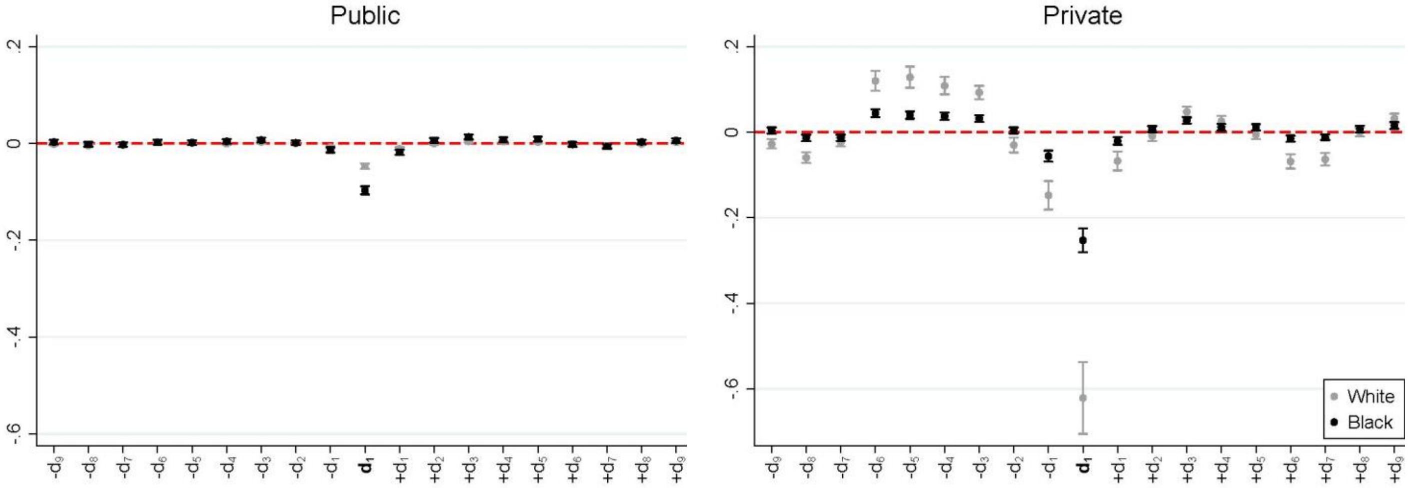
Panel C: Risk indicator



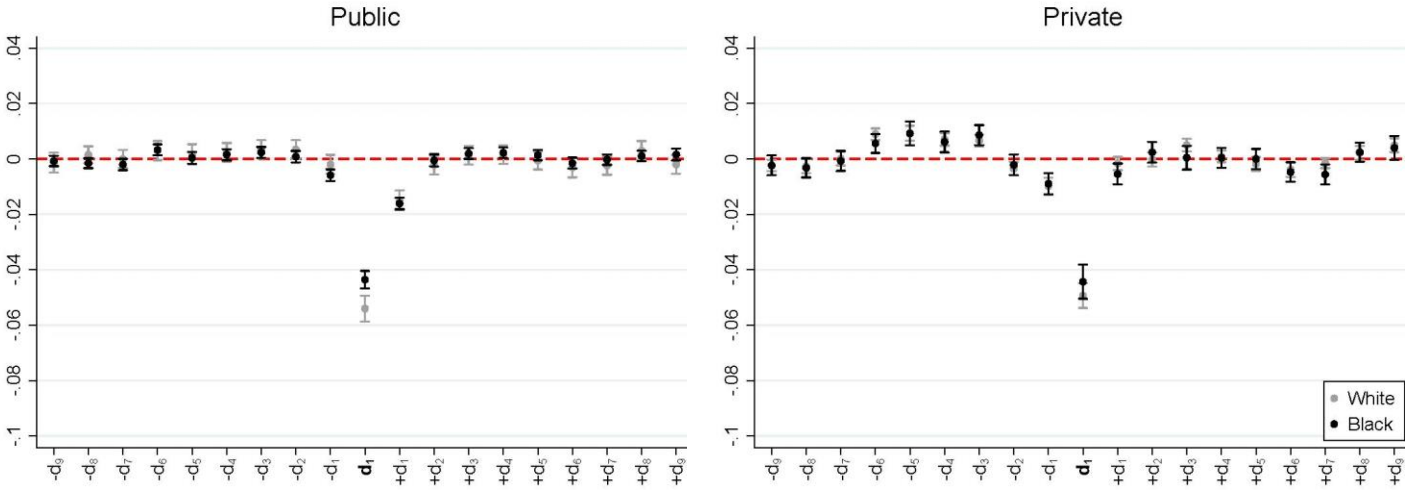
Notes: General plot description can be found in notes of Figure 5.1. The coefficient term -d_n (+d_n) refers to the n[th] day before the beginning (after the end) of the period encompassing *congress days*, represented by d_1, d_2, d_3, d_4, d_5. Estimate of coefficient d_5 is very imprecise as there is a single date in our dataset corresponding to the 5[th] congress day (i.e., all other congress events lasted 4 days). Whenever 95% CI exceeds our pre-defined scale, we replace it with arrows. Coefficients' point estimates and respective standard errors as well as the number of observations of each regression can be found in Table D.3.

Figure 5.3: Regression coefficients of days around *one-day bank holidays:* Number of births, delivery type, and risk profile

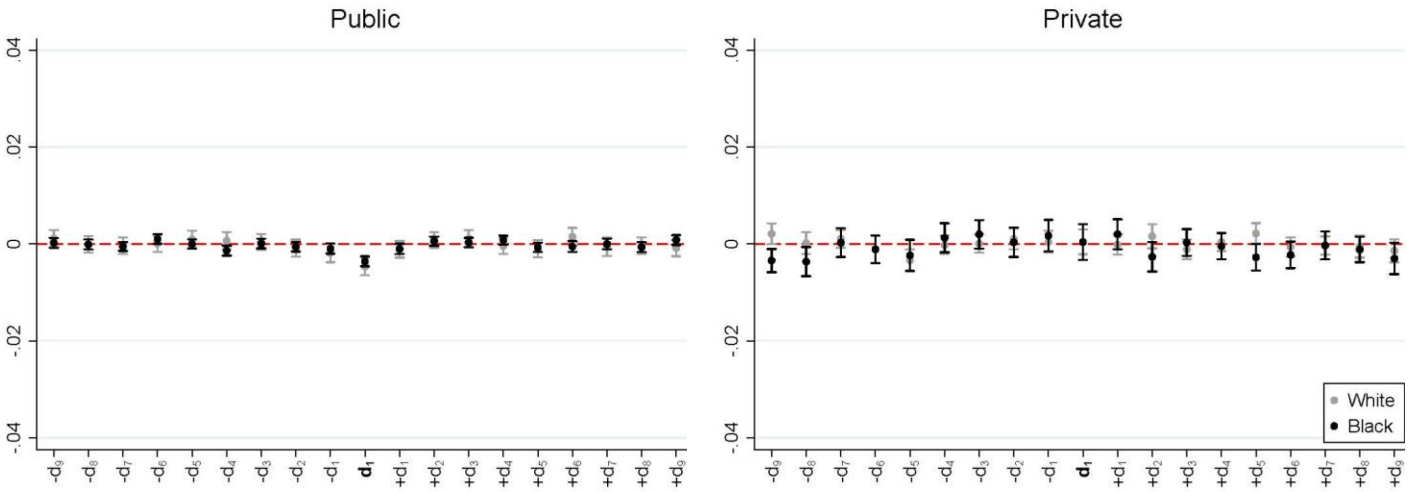
Panel A: Number of deliveries



Panel B: Likelihood of C-section



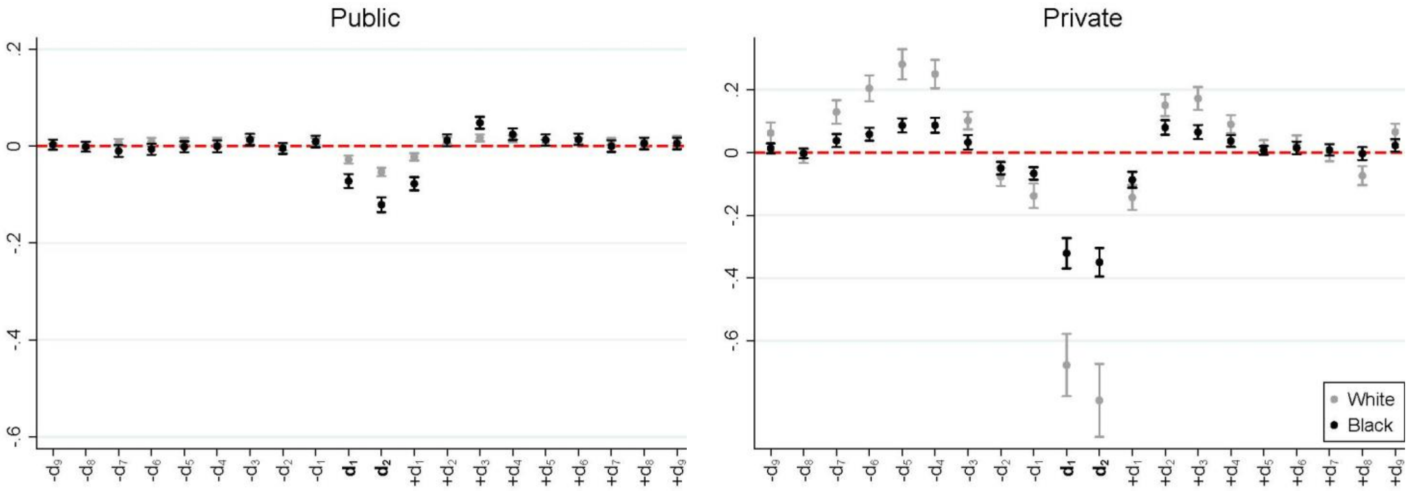
Panel C: Risk indicator



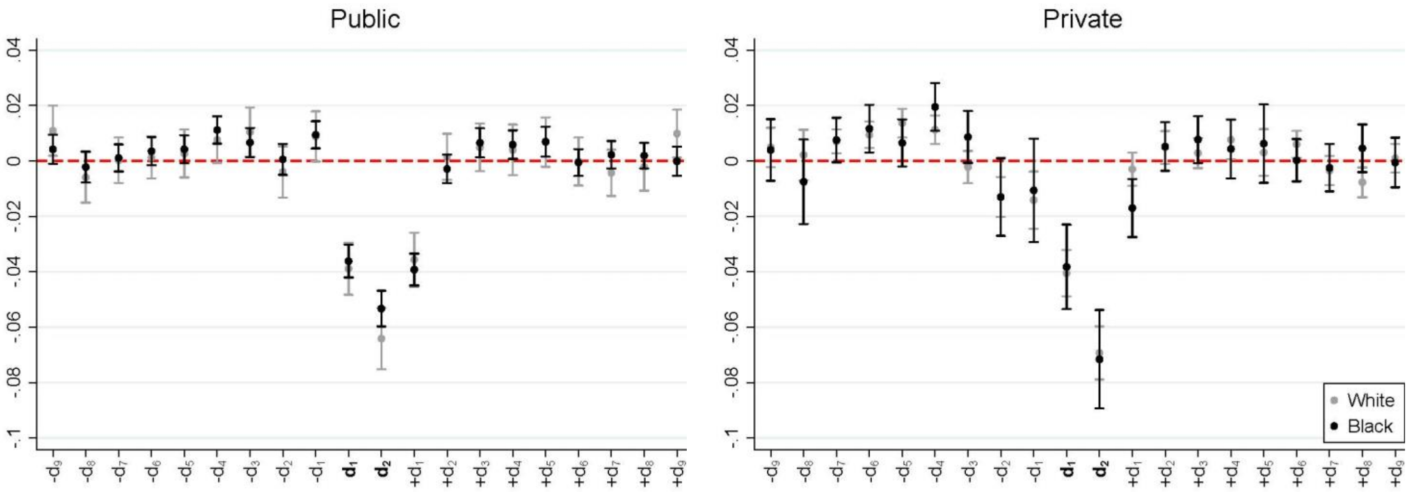
Notes: General plot description can be found in notes of Figure 5.1. The coefficient term -d_n (+d_n) refers to the n[th] day before (after) the *bank holiday*, represented by d_1. Coefficients' point estimates and respective standard errors as well as the number of observations of each regression can be found in Table D.4.

Figure 5.4: Regression coefficients of days around *two-day bank holidays*: Number of births, delivery type, and risk profile

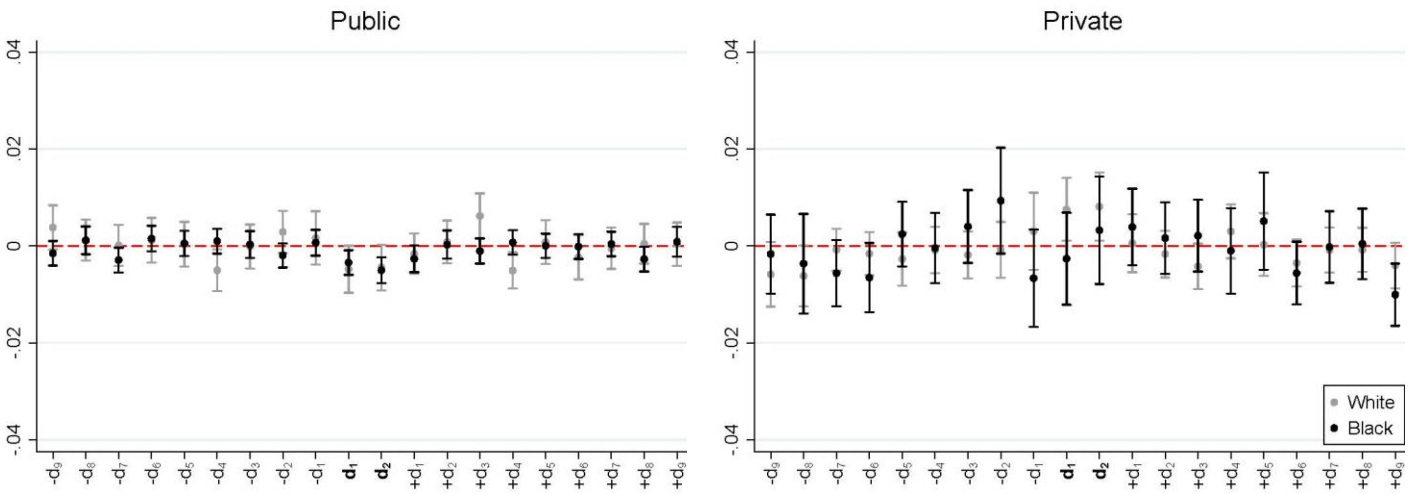
Panel A: Number of deliveries



Panel B: Likelihood of C-section



Panel C: Risk indicator



Notes: General plot description can be found in notes of .Figure 5.1. The coefficient term -d_n (+d_n) refers to the n[th] day before the beginning (after the end) of the period encompassing the *two-day bank holidays*, represented by d_1, d_2. Coefficients' point estimates and respective standard errors as well as the number of observations of each regression can be found in Table D.5

5.3.2 Racial gap in birth timing manipulation

To further investigate whether inequality stemming from birth timing manipulation is reduced in the public sector, we look at the difference in the number of births between racial groups as an outcome variable. In this way, we assess whether the racial gap varies across inconvenient and convenient days, in public versus private hospitals. We estimate regression model (5.1), with the dependent variable measured as number of deliveries by white mothers subtracted from number of deliveries by black mothers (i.e., number of excess births from black women). Figure 3.5 shows one coefficient plot for each inconvenience type, where each plot presents results of two different regressions: in blue are the estimates for the sample of births in public hospitals and in green those for the sample of deliveries in private hospitals.[149]

We find that, during inconvenient days, the number of births from black mothers increases relative to white mothers in private hospitals while no excess number of black births is seen in public hospitals. More specifically, we observe a decrease in the number of births from black relative to white mothers during bank holidays as well as congress days. In addition to being inconvenient, bank holidays are also days when risk of service delivery tends to be higher. Quality of available providers during congress days is also expected to decrease given that more skilled physicians are usually more likely to attend these very specialised conferences.

To check whether these results simply reflect the higher proportions of black women delivering in SUS and of white women in private hospitals, we estimate the same model for the share of births by black mothers as the outcome variable. Results, which we present in the appendix (Figure D.1), are in line with our previous argument. The bottom plots show that the proportion of births by black mothers during the days just before bank holidays tends to raise in SUS and drop in the private sector, which suggests that black (white) mothers are given priority within the public (private) sector in the manipulation of the timing of their births away from days when service delivery is expectedly lower.[150]

The results from the last section, indicating that birth timing manipulation in the public sector is primarily concentrated among instances where observed risk factors are more

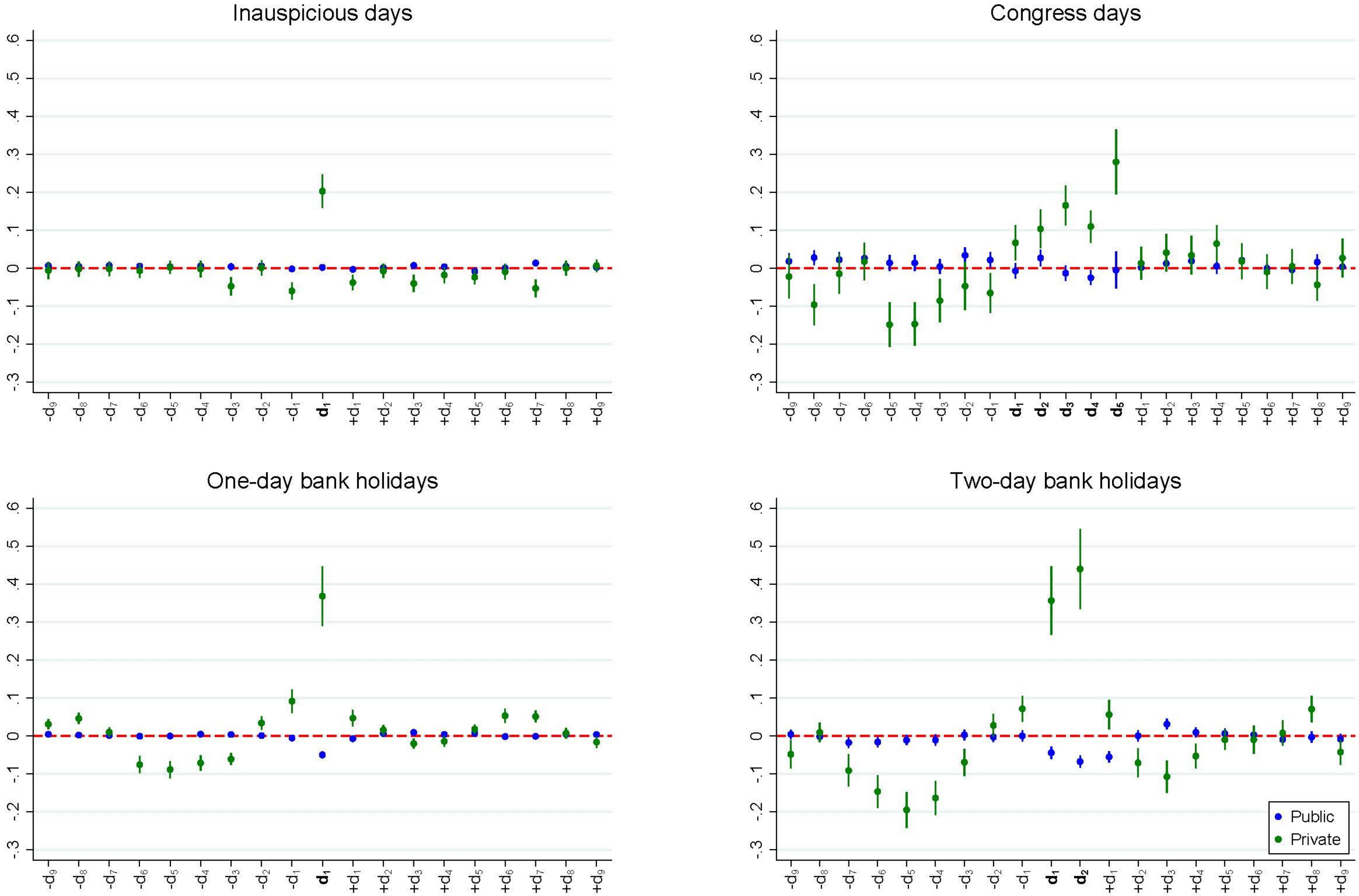
[149] Regression results (coefficients and standard errors) can be found in Table D.6.

[150] Although the coefficients are usually only statistically different from zero at the 10% level, they suggest clear trends.

prevalent, irrespective of the mother's race, imply that manipulation decisions are motivated by health concerns rather than skin color per se. The findings in this section highlight that, within public hospitals, black mothers are prioritized in the rescheduling of births away from days associated with overall higher procedural risks. If black mothers have higher underlying risk factors, many of which go largely unobserved, scheduling their births to occur when service quality is not compromised could serve as a protective measure against additional risks. This understanding is sensible under the assumption that the act of shifting the timing of births in this context does not meaningfully harm patient health, as suggested by available evidence in the literature.[151] Given that mitigating risk exposure among black mothers relative to white mothers is expected to alleviate underlying inequalities, this new finding reinforces the understanding that public hospitals serve as equalizers.

[151] Similar to our results for the public sector, Jacobson et al. (2021) find that manipulation of births away from holidays in California was concentrated among high-risk births. They find no health consequences from such manipulation, which are believed to be motivated by minimization of exposure to additional risks. Using data from a large public hospital in Italy, Fabbri et al. (2016) report similar dynamics of manipulation driven by risk aversion but do not assess health impacts. Evidence of negative health effects from birth timing manipulation have been largely restricted to contexts where the underlying motives are unrelated to concerns of patient health or health systems' constraints (e.g., parents' eligibility to financial rewards such as child tax benefits and baby bonuses).

Figure 5.5: Regression coefficients of days around inconvenient periods:
Excess births of black mothers



Notes: Outcome variable is the difference in the number of deliveries between black and white mothers. Unit level is hospital-day. Each plot shows point estimates and respective 95% confidence intervals for two separate regressions: one for births taking place in the Public sector (in blue) and another for those which happened in the Private sector (in green). Plot titles indicate the type of inconvenient day under consideration. Standard errors are clustered at the hospital level. For more details on regression specification, see Equation (5.1) in Section 5.2. The coefficient term -d_n (+d_n) refers to the n[th] day before the beginning (after the end) of the inconvenient period (displayed in bold). Coefficients' point estimates and respective standard errors as well as the number of observations of each regression can be found in Table D.6.

So far, we looked at hospitals which are exclusively public (i.e., 100% of obstetric beds affiliated to SUS during entire time sample) or exclusively private (i.e., 0% of obstetric beds affiliated to SUS during entire time sample). In a final analysis, we investigate whether this pattern is also observed as the same hospital becomes increasingly more (or less) associated with the public sector. We take advantage from the fact that hospital affiliation to SUS may change over time. Our data shows that 29% of the universe of 5,752 hospitals vary, throughout our time sample, in the reported monthly share of obstetric beds attached to SUS.[152] Among this sample of 1,654 hospitals, changes in SUS affiliation are substantial. The difference between the highest and lowest reported shares of these hospitals has an average of 39 and a median of 32 percentual points. The cumulative distribution is shown in the top plot of Figure D.2. The bottom plot, which draws the cumulative distribution for the overall variation throughout our time sample (i.e., share reported in first month by the given hospital subtracted from its reported share in the last month), shows that changes in hospital funding throughout our observed time horizon were reasonably symmetric across hospitals. At the median, we observe a decrease in the share of SUS obstetric beds of 5.8 percentual points throughout the analysis' time horizon.

To perform this last analysis, we estimate Equation (5.2),[153] using the entire sample of births (column 2 in Table 5.1).[154] Coefficients' point estimates and standard errors are presented in Table D.7. Based on these estimates, the figure below presents the total effect of inconvenient days by different levels of hospital affiliation to SUS.[155] Consistent with our hypothesis, as hospitals become more attached to public funding, the number of births from black relative to white women decreases during inconvenient periods, especially during

---

152 Among the universe of 5,752 hospitals, 2,900 hospitals present shares of SUS obstetric beds fixed at 100% (i.e., exclusively public) while 875 hospitals present shares of SUS obstetric beds fixed at 0% (i.e., exclusively private) as reported in Table 5.1. There are 323 facilities who report fixed shares at values other than 0% or 100%. The remaining 1,654 hospitals present varying shares over the months of our analysis' time sample. This is the sample we refer to in this section.

153 In Equation (5.2), we regress the difference in the number of births delivered by black relative to white mothers at the hospital-day level on indicator variables of the share of SUS obstetric beds reported by the given hospital in the given month, dummy indicator of inconvenient periods, dummy indicator of days in the vicinity of inconvenient periods, and interaction terms between the first and the last two. Time fixed effect (year, month, and day of the week) and hospital-specific effects are also included. Standard errors are clustered at the hospital level.

154 Although identification comes from hospitals for which exposure to SUS vary throughout our time sample, our estimation includes the entire sample of births in order to gain precision.

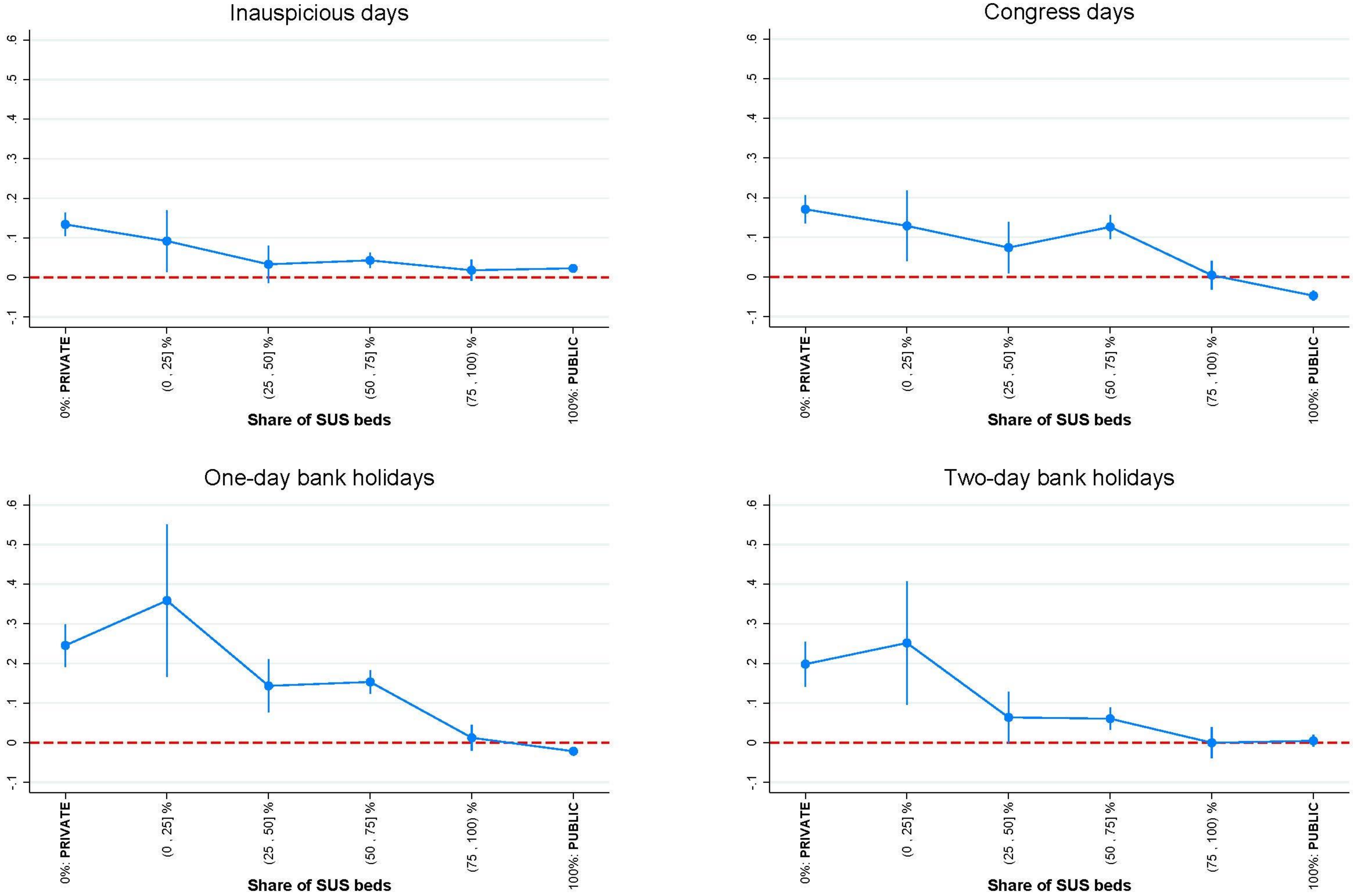
155 This is computed as $\hat{\kappa} + \hat{\beta}_r$, where each term represents the OLS estimate of the respective parameter in Equation (5.2), and the index r (which is represented in the plots' horizontal axis) informs the degree of affiliation to SUS by the given hospital in the given month.

congress periods and bank holidays, where hospital's service delivery faces higher procedural risks.

Identification comes from a distinct sample of hospitals – those with varying affiliations to SUS over the months of our time sample. Previous findings, in contrast, are based on hospitals reporting fixed shares of SUS obstetric beds at either 0% or 100% throughout all reported months. It is reassuring that the dynamics previously identified by comparing exclusively public to exclusively private hospitals are also observed within hospitals as they transition between serving the two health systems. This, again, reinforce our understanding that public health system alleviates risk exposure for childbirths which are likely to be most affected by them. This is expected to be welfare-improving as long as manipulating the timing of births does not introduce risks to individual births which are greater in magnitude to those offset by the expected lower procedural risks present at the actual time they take place.[156]

[156] Indeed, current literature suggests that patient health is likely to be largely unaffected by birth timing manipulation in settings like ours (Jacobson et al., 2021). Negative health consequences previously documented in the literature have been limited to contexts where manipulation is motivated by rewards which are solely extracted by parents if births are delivered by a given predetermined date (e.g., tax benefits, baby bonuses). For more details, see footnote 151.

Figure 5.6: Regression coefficients of inconvenient periods, by SUS affiliation: Excess births of black mothers



Notes: This figure outputs the effect of inconvenient days on the gap between the number of births by black mothers relative to the number of births by white mothers taking place in hospitals which present the share of SUS obstetric beds in the given month as informed in the x-axis. Each plot shows both point estimates and respective 95% confidence intervals of such effect for a given type of inconvenient day (informed in the title of the plot). Estimates presented in each plot come from the same estimated model, according to Equation (5.2). The point estimates plotted above refer to the sum of the regression's coefficients of inconvenient day and its interaction term with the given interval corresponding to the share of SUS obstetric beds (i.e., estimates of parameters κ and β_r in Equation (5.2). Unit level is hospital-day. Standard errors are clustered at the hospital level. Coefficients' point estimates and respective standard errors as well as the number of observations of each regression can be found in Table D.7.

5.4 Discussion

In this chapter we examined birth timing by using data on over 20 million births in Brazil with the goal to document patterns of manipulation around different inconvenient periods. These patterns are studied by race of the mother as well as across health systems.

Using hospital-day level data, we rely on very granular variation to investigate the distribution of births, mode of delivery, and risk profile around inconvenient days. In a fixed effect regression model, we are able to account for heterogeneity of hospitals as well as month, year, and weekday when births are delivered.

Estimates would be biased if birth records suffer from measurement errors or misreporting of observables, such as mothers' race and risk factors. This would only be a concern if these mismeasurements and misreporting are systematically distributed across inconvenient periods and their vicinities. Additionally, the likelihood of misreporting during inconvenient periods is mitigated by the fact that birth information is typically recorded not on the exact date of childbirth, but rather upon discharge from the hospital, which may occur several days later. If miscoding is not systematically distributed across inconvenient periods, we should anticipate larger standard errors. Our sample size is sufficiently large for precise inference within clusters.

We found different patterns of birth timing between white and black mothers, thus revealing inequalities, and characterized how these patterns behave within different hospital systems. While manipulation generally occurs around different inconvenient periods, it is greater in the private sector and among white women. The results therefore suggest the existence of choice flexibility in the private sector, which allows manipulation of the timing of birth to manifest due to convenience reasons from both sides, parents and physicians, as indicated by delivery patterns around inauspicious and congress days. Convenience reasons from both sides are also present in birth timing around bank holidays in private hospitals, as substantial manipulation occurs without significant changes in the risk profile of births.

On the other hand, manipulation is more limited in the public sector while differences between white and black women are largely equalized, as indicated by statistically similar patterns around inauspicious and congress days. Also, within the public sector, we found relatively more manipulation on bank holidays in comparison to other inconvenient periods, but manipulation is accompanied by changes in the risk profile of births -- the likelihood of riskier deliveries is reduced during bank holidays, and this result is again similar both for white and black women. This finding suggests that some manipulation occurs because of medical appropriateness aimed at minimizing exposure to procedural risks in times of resource scarcity. Whereas the number of excess births from black women decreases in public hospitals during bank holidays and congress days (when quality of available doctor is expectedly lower), we observe very marked patterns in the opposite direction among private facilities. This same pattern is found for a separate sample of facilities experiencing varying exposure to the public sector over the time period of our analysis: as their affiliation to SUS increased (decreased), manipulation away from days where risk is likely higher became more (less) targeted at black mothers.

In addition to being inconvenient times to work and deliver, bank holidays are also times when risk of being admitted to hospitals is likely higher due to lower availability of hospital staff, lower levels of experience of medical team on shift as well as potentially lower level of support and supervision which could, in turn, affect staff concentration and performance. There is a large number of studies pointing to a negative association between hospital admissions during off-hour periods (i.e., bank holidays, weekends, evening hours) and health outcomes in case of childbirth (Gould et al., 2003; Hong et al., 2006; Palmer et al., 2015) as well as other procedures (Becker, 2007; Magid et al., 2005; Zapf et al., 2015). The literature suggests that variations in health outcomes between regular business hours and off-hours are not fully explained by differences in the case-mix, and that differences in organizational factors may be an important contributor (Becker, 2007; Hamilton et al., 2007; Hong et al., 2006). Indeed, there is documentation of inferior nurse-to-patient ratio and lower compliance to standardized protocols and guidelines in Brazilian ICUs during leisure days (Zampieri et al., 2018). Most evidence highlighting the negative association between time of admission and health outcomes in Brazil has focused on emergency hospital admissions due to heart attack (Barros et al., 2013; Evangelista et al., 2008; Leivas, 2017), where longer times to treatment seems to be one of the main reasons explaining worse outcomes during off-hour periods (Becker, 2007; Magid et al., 2005). If waiting times are also longer for women who go into labour during off-hours, there would be higher scope for emergency C-sections to go undetected in case labour does not go as expected.

The primary limitation of this study lies in the absence of health-related outcomes, which impedes our ability to ascertain the consequences of manipulating birth timing. Another constraint of this study is that we are not able to track mothers during the gestational period up to delivery and we cannot identify whether they change the location of delivery (public *vs.* private facilities) in a way that is systematically correlated with inconvenient periods. We conjecture that leakage between sectors led by rescheduling away from very specific inconvenient periods should not be a concern as the exact date of delivery is not accurately predicted ex-ante.

Most of the literature examining racial disparities during childbirth has centred on procedure appropriateness. For instance, Valdes (2021) document that black and Asian women in the US exhibit higher C-section rates in the subset of low-risk births, where medically justified C-sections are less common, and lower rates in the subset of high-risk births, where C-sections are expected to be largely beneficial. Robinson et al. (2023) show that black American mothers with the highest measured C-section appropriateness receive C-sections less frequently than similarly-appropriate white mothers. To the best of our knowledge, this study is the first to explore racial gaps in the timing of childbirth procedures. Furthermore, it contributes to the existing literature by scrutinizing how the racial gradient varies across different healthcare systems within the same country.

Our findings suggest that birth timing manipulation in the public sector is mostly motivated by risk aversion and equity concerns. This is not surprising given its ambition in providing universal access free at the point of use for everyone as well as its organisation with incentives uniformly distributed (i.e., not determined by type or duration of procedures). Although public hospitals tend to face higher-order constraints, it also seems to have the capacity to offset potential risks to patients most likely to benefit from it in times when resources are scarcer and processes suboptimal. First, we find that, in times when patient admission is generally riskier, manipulation is concentrated among births that are subject to more critical (observed) risk factors. Second, we show that manipulation away from these times is relatively more widespread among black women, who traditionally come from lower socioeconomic backgrounds and experience greater underlying health vulnerabilities. Given the high prevalence of C-sections in Brazil's private sector, where altering timing typically doesn't impact the delivery method, scheduling births for convenience reasons is commonly expected. If timing deliveries due to non-medical factors have negative health consequences to mothers or their babies, those opting for an unregulated sector could be subjected to unnecessary risk.

6 FINAL REMARKS

This thesis focuses on examining different factors influencing physicians' decision-making in medical treatment. Clinical decisions are shaped by a multitude of supply-side factors, encompassing the accuracy of physicians' diagnoses, their awareness of scientific evidence regarding treatment efficacy, and their skills in executing treatment procedures. Moreover, inherent uncertainty in medical knowledge makes room for varying opinions on treatment appropriateness, especially given differences in educational background, firsthand experience (e.g., pool of treated patients, returns to specialisation), and individual preferences. Additionally, medical decisions are subject to the influence of the clinical practice environment and institutional incentives or constraints. Furthermore, non-medical factors, such as financial incentives, convenience considerations, and willingness to accommodate patient preferences, can also play a role in shaping clinical choices.

Our first empirical chapter shows that peers play an important role in determining physicians' treatment choices. It specifically looks at the overall costs of physicians' medical decisions to the public health system in Brazil. First, we show that being recently exposed to higher shares of female peers causes physicians to reduce their medical spending. This is particularly relevant given recent trends of increasing female representation in medical occupations. Second, we find that doctors who are exposed to more (less) resource-intensive physicians become more (less) resource-intensive themselves. While some preferences are harder to change, peer interaction provides physicians with the opportunity to learn from colleagues with different educational backgrounds and accumulated experience. Besides, physicians who are driven to experiment alternative treatments are likely to gain new procedural skills. Alternatively, physicians may increase their spending if surrounded by less resource-responsible peers due to perceived lower accountability. This could happen if, for instance, physicians become more likely to order unnecessary tests or provide care that is more intensive than needed.

Although the use of more expensive treatment alternatives that do not add value to the provision of care is usually more common in health systems where financial incentives encourage providers' induced demand, this could also happen if there are other incentives motivating its provision, such as convenience and fear of litigation, or in case of no clear evidence of treatment returns to patient health. Indeed, we show evidence in this direction in the context of C-section choice during childbirth delivery. The following two analytical chapters document evidence on widespread use of unnecessary C-sections as well as the role of convenience motivations during childbirth care, especially in the private sector.

We begin by documenting substantial evidence of variation in C-section rate across Brazilian municipalities. While geographical variation in treatment patterns is per se not necessarily suboptimal (i.e., if justified by medical uncertainty, knowledge spillovers, or physicians' comparative advantage), we present compelling evidence that it is when it comes to the large observed discrepancies in the propensity to perform C-section in Brazil. We assess the effects from a federal reform that restricted the relative use of C-sections by introducing a compensation cap to the share of C-sections delivered monthly in hospitals attached to the public health system. Our results show that the reform was successful in reducing C-section likelihood in municipalities with higher baseline propensity to perform this delivery method. The fact that the policy significantly decreased C-section use, as intended, while improving infant health is highly informative: not only unnecessary C-sections were prevalent prior to the reform, but also those which were detrimental to patient health. We argue that the negative consequences of medically unjustified C-section are still not broadly recognised although recent research (to which we contribute) increasingly points in this direction.

Our final empirical investigation presents evidence that both mothers' and physicians' convenience motivations influence the timing that births are delivered, especially in the private sector. We provide evidence that the timing of births is manipulated to occur away from inauspicious days (inconvenient for mothers) and away from days when the Brazilian Congress of Obstetricians and Gynaecologists is held (inconvenient for physicians). Deliveries are also systematically shifted away from bank holidays, which in addition to being characterised by inconvenience are also times when risk of hospital service is higher because of scarcer resources available. While manipulation away from bank holidays is more substantial and happens independently of the risk profile of births in private hospitals, manipulation is targeted at high-risk births within the public sector. We read the latter as evidence of attempts to offset potential risks to patients in more vulnerable health states. Furthermore, birth timing manipulation driven by convenience motives is likely to be more prevalent in the private sector because changes in delivery method are less often required given that C-section use (which allows manipulation of the timing of birth) is much more frequent in this setting.

Summing up, we find that physician decision-making is, to some extent, *malleable*. Physician behaviour is largely influenced by the clinical environment where they practice (i.e., exposure to peers of different characteristics and treatment styles) as well as by the institutional incentives in place (i.e., hospital-level financial incentives). Additionally, motivations behind medical decisions differ considerably across health systems. We show that while in the private sector medical choices are influenced more significantly by non-medical incentives of both patients and physicians (i.e., patient preferences, physicians' demand for leisure), decisions on how to allocate care in the public health system seems to occur partly in response to patient risk factors. This is driven, on the one hand, by different features of each system (i.e., higher flexibility in treatment choice in the private sector, organisation of public system around disease severity) and, on the other hand, selection of physicians across health systems based on their individual preferences (i.e., more altruistic physicians selecting into public hospitals).

One of the main strengths of the thesis lies in its extensive use of large datasets. While Chapters 4 and 5 draw upon comprehensive birth record data, Chapter 3 uses information of all procedures conducted within Brazil's public healthcare system. Moreover, we are able to pinpoint the specific physician responsible for each procedure, for which we also possess detailed demographic characteristics, such as date of birth, gender, university of medical degree, and indicator of completion of medical residency. Through this analysis, we can meticulously track physicians' activities within the public sector over a span of 7.5 years and discern whether they also engage in private sector services, albeit without direct observation of this sector's activities. In Chapter 4, our analysis incorporates data from both public and private sectors, yielding pertinent local average estimates, yet without differentiation between the two. On the other hand, the analysis in Chapter 5 allows us to examine variations in physicians' activities between the public and private sectors - a pivotal aspect within Brazil's healthcare system, characterized by universal coverage free at the point of use alongside a robust private sector presence. Finally, another crucial advantage of the studies in this thesis is their focus on a middle-income country. The majority of research concerning variations in medical practice originates from high-income countries, such as the US and European nations. It is, however, particularly imperative to find more efficient ways to allocate available resources in the developing world, where health systems face severe constraints. Redirecting resources to areas of care with greater need and higher health returns holds immense potential to substantially enhance patient welfare in these settings.

The evidence laid out in this thesis carries profound implications, especially considering that a significant portion of the disparity in treatment approaches among healthcare providers stems from less-than-ideal medical judgments. This underscores the considerable opportunity for policymakers to enact interventions and drive improvement in healthcare decision-making. Our findings posit that interventions could include policies aimed at influencing peer composition, promoting knowledge exchange and collaboration among team members, implementing changes to institutional incentives and constraints, and increasing regulation of the private sector when deemed necessary.

We emphasize the need for further research to advance the understanding of how guidelines can support physicians in making increasingly optimal treatment decisions. While guidelines could be beneficial to enhance the quality of decision-making in areas where current medical knowledge points to a clear consensus, they could also prove helpful for highly complex decisions. A promising area of research involves the development of recommendation tools that more effectively incorporate uncertainty in expected success rates of treatment options by patient type, allowing physicians the flexibility to choose from undominated treatment alternatives. These guideline tools could also assist physicians in evaluating patients' risk factors and underlying medical conditions by carefully considering clinical traits and observed symptoms. In recent work, Manski discusses how guidelines could help physicians in achieving optimal clinical decisions, given available information and deep uncertainty (Manski, 2017, 2018, 2019). Research should additionally aid policymakers and insurers in exploring alternative measures to address the wide variation in physicians' diagnostic skills. This could include initiatives like implementing diagnostic training or restructuring health systems to enhance diagnosis accuracy. Such efforts may involve validating complex individual decisions through peer input and structuring separate teams for diagnostic assessments and treatment selection, acknowledging the distinct skill sets required for each task.

7 APPENDICES

Appendix A summarises existing evidence on physicians' practice styles which adds to Chapter 2. Appendices B and C presents additional tables and figures to Chapter 3 and Chapter 5, respectively.

List of Appendix Tables (Chapters 3 & 5)

List of Appendix Figures (Chapters 3 & 5)

APPENDIX A: EVIDENCE OF PHYSICIANS' PRACTICE STYLES (CHAPTER 2)

The main challenge in attributing variation in treatment dynamics to individual physicians is that the assignment of patients to physicians is usually endogenous. Some physicians may provide more intensive care simply because they see patients who are sicker (and would consequently benefit more from these services) and/or patients who have higher preferences for such procedures (and consequently are likely to overemphasise the severity of their symptoms and/or self-select to be seen by physicians who specialized in the type of procedure they would like to receive). Because available data doesn't inform patient preferences and all relevant risk factors, researchers need to find reliable strategies to be reassured that findings are not driven by these confoundeness.

The most straightforward solution is to look for institutional settings where patient-physician sorting is limited because of the way processes are designed. The safest bet typically is the hospital's emergency department, where care to patients is usually provided on a first come first serve basis and urgency of health condition prevents patient sorting across hospitals. Van Parys (2016) looks at variation within emergency departments in the US and find that physicians at the 75th percentile of the spending distribution consume 20% more resources than physicians at the 25th percentile. Using Medicare data of patients hospitalized with a nonelective medical condition, Tsugawa et al. (2017) show that physicians in the highest-utilization quartile have an adjusted average spending 40% higher than those in the lowest quartile.

Other studies attempted to limit physician-patient endogenous matching by saturating the model with many control variables. The underlying assumption for causal inference is that, conditional on observables, there is no systematic variation in the distribution of patients' underlying health across physicians. With Nordish data, Grytten and Sørensen (2003) estimate a model of primary healthcare use as a function of patient characteristics and fixed effects of region, time, and physician to then compare its explained variation with that of another model which does not include physician fixed effect. They find that physician-specific factors are responsible for ~50% of the total variation in healthcare spending. Epstein and Nicholson (2009) estimate related models for obstetricians in the US and show that the variation of risk-adjusted C-section rate across physician in the same region is 75% higher than the average risk-adjusted rate across regions.

Researchers have also turned to settings where patients are typically treated by a single physician (i.e., assigned primary care physician / GP) but experience switching to another physician over time. The underlying assumption for identification is that patient mobility across providers is conditionally exogenous (i.e., not related to health seeking behaviour) - usually a more reasonable assumption when patient-physician separation is supply-driven. Fadlon and Van Parys (2020) exploit variation spurred by physician exit from the local healthcare system due to migration or retirement.[157] They find an immediate and long-lasting impact in the healthcare utilization of patients who had to switch providers. This is measured in a number of ways such as health spending, number of doctor visits, number of diagnoses received, probability of guideline-consistent care, and avoidable hospitalizations. Kwok (2019) and Ahammer and Schober (2020) also exploit mobility between patients and their assigned primary care physicians in the US and Austria, respectively.[158] Kwok finds that physician-specific factors explain, on average, 13% of the within-region variation of adjusted health utilization in the long run (net of temporal switching effect). Ahammer and Schober show that physicians in the top decile present an average healthcare utilization that is 25% higher than the average physician. Using similar variation in Danish population-level data, Huang and Ullrich (2023) present findings that practice style heterogeneity accounts for more than half of the variation in total antibiotic prescribing behaviour across physicians in primary care.[159,160]

[157] Although separations from old physicians are expected to be exogenous as they are triggered by unilateral physician decisions, patient choice of new physician could be endogenous. The identification assumption is that patients' healthcare utilization after physician exit is parallel to what it would have been if their physician would not have left the health system.

[158] If time of switch is determined by health shocks, for instance, the identifying assumption would be violated. Evidence of no increase of healthcare use prior to switch is reassuring but does not guarantee that time of switch is exogenous from changes in underlying health.

[159] Earlier studies by Phelps estimate this figure to be between 25% and 60% for primary care physicians in the US. His econometric framework is less robust as they only control for severity of illness and observed patients' characteristics (Phelps, 2000; Phelps et al., 1994).

[160] The commonly adopted methodologies are: (i) differences-in-differences specifications with patient fixed effect, where the main regressor of interest is an interaction between the switch indicator and differences in average healthcare usage between the patient's new and old physicians. The coefficient of the interaction term is interpreted as the share of this utilization gap between providers that can be explained by differences in their practice styles; and (ii) two-way fixed effect models, with fixed effects of patient and physician, followed by a decomposition analysis that computes the share of physician-specific contribution. These papers rely on variation generated from different timing of switches, which recent literature has shown to be problematic in case of heterogeneous effects (Goodman-Bacon, 2021). Because most of these papers predate the current understanding of most appropriate econometric methods to be used in these cases, findings should be taken with some caution. Recently, Huang and Ullrich (2023) showed evidence that their results are robust to these new methods – a reassuring finding.

Tu (2017) goes a step beyond by exploiting variation across specialist doctors treating patients diagnosed with the same medical condition by the same primary care physician. In his job market paper, the author uses changes to the referral networks of Medicare primary care physicians triggered by the exit of a specialist. If care is entirely determined by patient characteristics, then changing the set of specialists should not alter care received. If, instead, variation in care is fully driven by physician-specific factors, then the utilization should change to reflect the practice style of the new network of specialist physicians. The author finds that 50-70% of variation in health

care usage is driven by doctor styles and that higher initial utilization leads to greater subsequent utilization. Causal interpretation relies on the assumptions that neither patients nor primary care physicians directly respond to the network change by, respectively, switching primary care provider or adjusting care patterns.

APPENDIX B: ADDITIONAL TABLES AND FIGURES (CHAPTER 3)

Table B.1: Descriptive statistics by range values of the instrument

	All	By interval of % females among Peers of Peers (IV)					
	[0, 1]	[0, 0.2[	[0.000, 0.025[	[0.025, 0.075[	[0.075, 0.125[	[0.125, 0.175[	[0.175, 0.225[
N. obs	13,502,212	5,090,839	1,779,339	642,819	885,707	1,182,034	1,364,186
% obs	100%	38%	13%	5%	7%	9%	10%
Municipality							
State capital	48%	43%	28%	45%	51%	51%	52%
South/Southeast regions	48%	40%	30%	37%	42%	48%	54%
Health facility							
General hospital	83%	84%	86%	85%	81%	84%	82%
Specialised hospital	15%	13%	12%	11%	14%	13%	15%
Teaching status	45%	41%	26%	42%	48%	49%	52%
General admission	84%	83%	82%	82%	83%	84%	85%
Medical specialty							
Focal physician (most frequently reported)							
General medicine	32%	35%	40%	36%	34%	32%	30%
Obstetrics/Gynaecology	26%	28%	28%	29%	28%	27%	27%
General surgery	14%	10%	11%	11%	9%	9%	11%
Orthopaedics	7%	7%	7%	6%	6%	7%	7%
Paediatrics	8%	8%	5%	7%	9%	11%	10%
Shared Focal - Peers (most common)							
General medicine	45%	51%	58%	55%	47%	44%	41%
Obstetrics/Gynaecology	21%	21%	19%	20%	23%	22%	23%
General surgery	14%	9%	9%	9%	9%	9%	11%
Orthopaedics	7%	6%	6%	5%	5%	6%	6%
Paediatrics	6%	6%	3%	5%	7%	9%	8%
Shared Peers - Peers of Peers (most common)							
General medicine	34%	26%	16%	24%	30%	32%	37%
Obstetrics/Gynaecology	9%	5%	7%	3%	4%	4%	5%
General surgery	36%	46%	39%	46%	49%	50%	46%
Orthopaedics	5%	2%	1%	2%	2%	2%	2%
Paediatrics	6%	11%	18%	17%	8%	5%	3%

Notes: The table provides statistics for values of the instrumental variable within the specified interval bracket in each column's title. The first column considers all observations in our estimation sample. The second column displays statistics for the subsample where the share of females among peers of peers is below 20%. Subsequent columns further break down this interval into smaller brackets, aligning precisely with those depicted on the x-axis of Figure 3.3.

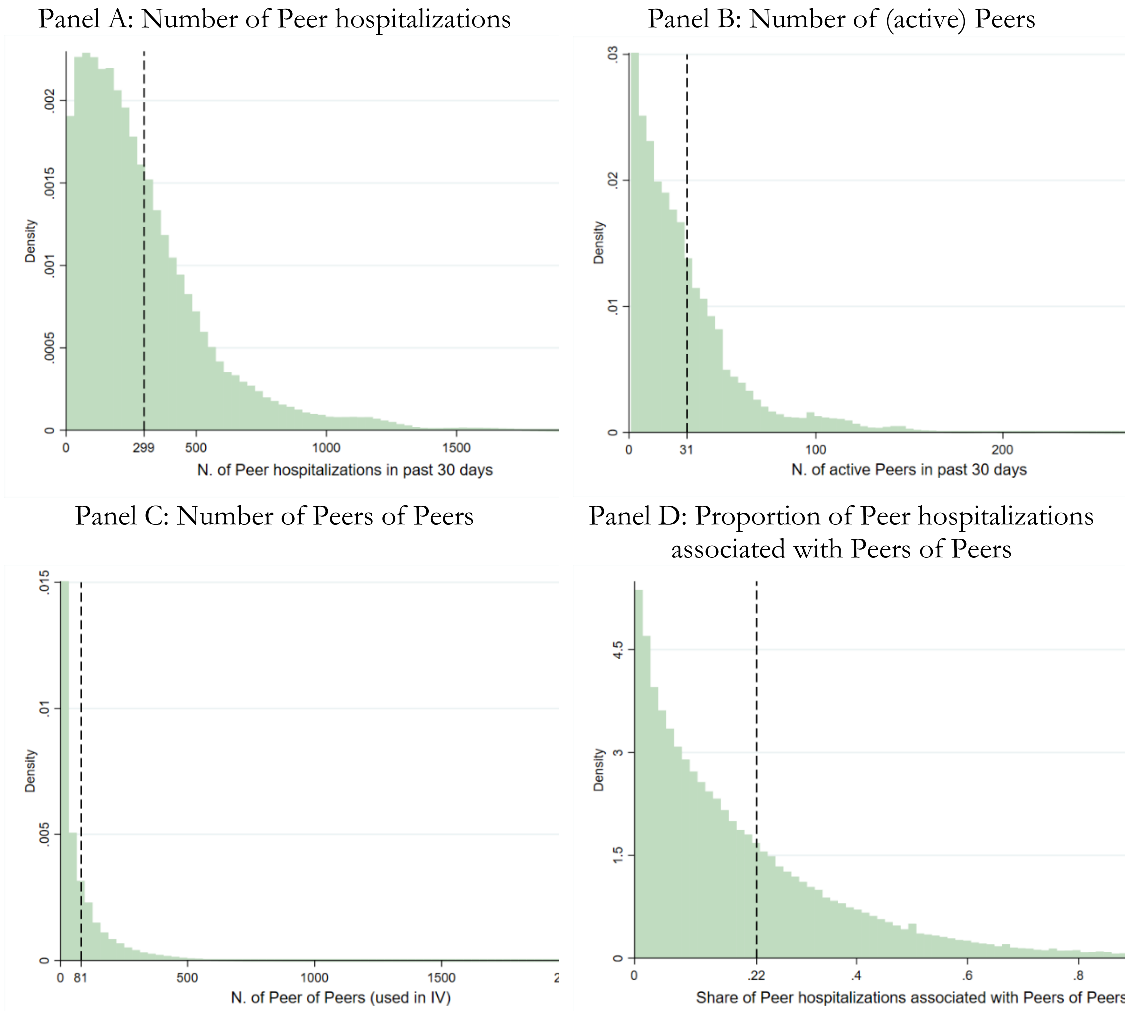
Table B.2: Efficient two-step GMM estimates: different parametric specifications

Outcome:	Ln of Hospitalization Cost				
IV:	Linear	Linear pw knot: 0.2	Linear pw knots: 0.25 & 0.75	Quadratic	Cubic
Peers' average outcome	0.206	0.595***	0.614***	0.526***	0.574***
	(0.171)	(0.100)	(0.092)	(0.110)	(0.092)
Peers' characteristics					
% female	-0.374***			-0.644***	-0.756***
	(0.082)			(0.114)	(0.156)
% female (0.00 - 0.20)		-0.609***			
		(0.113)			
% female (0.20 - 1.00)		-0.123***			
		(0.046)			
% female (0.00 - 0.25)			-0.522***		
			(0.091)		
% female (0.25 - 0.75)			-0.108**		
			(0.054)		
% female (0.75 - 1.00)			-0.026		
			(0.093)		
% female ^ 2				0.477***	0.925***
				(0.098)	(0.341)
% female ^ 3					-0.359
					(0.221)
average age	0.002	0.001	0.001	-0.005	-0.035
	(0.002)	(0.001)	(0.001)	(0.006)	(0.031)
average age ^ 2				0.000	0.001
				(0.000)	(0.001)
average age ^ 3					-0.000
					(0.000)
% top university degree	-0.035	-0.012	-0.011	0.088*	0.164
	(0.035)	(0.024)	(0.023)	(0.049)	(0.116)
% top university degree ^ 2				-0.122***	-0.359
				(0.047)	(0.349)
% top university degree ^ 3					0.178
					(0.244)
% residency degree	0.205***	0.115***	0.122***	0.316***	-0.066
	(0.047)	(0.031)	(0.033)	(0.074)	(0.167)
% residency degree ^ 2				-0.177***	0.684*
				(0.068)	(0.376)
% residency degree ^ 3					-0.547**
					(0.258)
Focal's characteristics					
female	-0.069***	-0.058***	-0.054***	-0.058***	-0.057***
	(0.008)	(0.010)	(0.009)	(0.008)	(0.008)
age	0.001**	0.000	0.000	0.003	0.019*
	(0.000)	(0.000)	(0.000)	(0.002)	(0.011)
age ^ 2				-0.000	-0.000
				(0.000)	(0.000)
age ^ 3					0.000
					(0.000)
top university degree	-0.016**	-0.016**	-0.016**	-0.016***	-0.016***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
residency degree	0.056***	0.068***	0.071***	0.062***	0.056***
	(0.021)	(0.018)	(0.017)	(0.018)	(0.016)
residency degree, ever	0.033	0.014	0.012	0.024	0.026*
	(0.020)	(0.017)	(0.016)	(0.017)	(0.015)
multiple specialty	0.074***	0.077***	0.076***	0.077***	0.076***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
employment: autonomous	0.055***	0.041***	0.037***	0.036**	0.038***
(vs staff)	(0.020)	(0.012)	(0.012)	(0.014)	(0.012)
employment: other	0.116**	0.099***	0.098***	0.119***	0.111***
(vs staff)	(0.049)	(0.034)	(0.034)	(0.038)	(0.034)

Outcome:	Ln of Hospitalization Cost				
IV:	Linear	Linear pw knot: 0.2	Linear pw knots: 0.25 & 0.75	Quadratic	Cubic
Patient					
female	-0.034***	-0.025***	-0.025***	-0.026***	-0.025***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
age	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Health facility					
teaching status	0.145**	0.049*	0.046*	0.070*	0.056**
	(0.057)	(0.028)	(0.026)	(0.037)	(0.028)
general admission protocol	0.055*	0.048**	0.052**	0.060**	0.052**
	(0.033)	(0.024)	(0.022)	(0.025)	(0.021)
type: specialised	0.028	0.036*	0.034*	0.029	0.031
(vs general hospital)	(0.025)	(0.019)	(0.019)	(0.020)	(0.019)
type: other	0.065	-0.013	-0.010	0.021	-0.004
(vs general hospital)	(0.106)	(0.063)	(0.062)	(0.076)	(0.065)
Diagnostic group FE	X	X	X	X	X
Municipality FE	X	X	X	X	X
Year/month FE	X	X	X	X	X
Observations	13,502,212	13,502,212	13,502,212	13,502,212	13,502,212

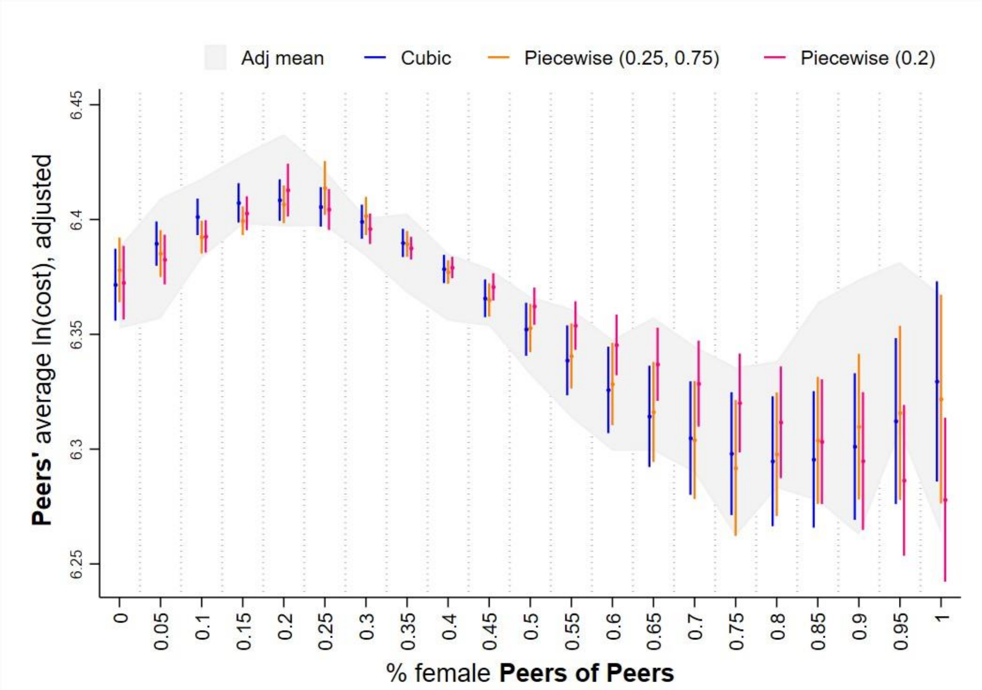
Notes: Peer's average outcome is instrumented with share of females among peers of peers. For consistency, in higher order polynomial specifications (last two columns), we modelled other physician characteristics in the same way as gender. Results are very similar when we model these variables linearly. For linear piecewise regressions, we restrict the knot(s) solely to physician gender. This is because the exact knots were specifically defined for this variable (i.e., female representation), as pointed both by our first-stage non-parametric relationship investigation (as seen in Figure 3.3). Standard errors are clustered at the municipality level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure B.1: Distribution of number of peer hospitalizations, peers, and peers of peers



Notes: Each observation refers to our estimation sample (described in Section 3.4.4). Panel A presents number of peer hospitalizations considered in the peer outcome variable in the right-hand side of our regression model. Panel B shows number of peers who led such hospitalizations. Panel C informs the number of peers of these peers in the relevant calendar month, which are used in our IV construction. Peers of peers are unrelated to focal physician in both hospital and specialty dimensions. The vertical dashed line depicts the average value, across all observations in our sample, of the respective variable whose distribution is being plotted.

Figure B.2: Predicted peer outcome by IV: cubic vs linear piecewise specifications



Notes: The 95% confidence interval, displayed in light grey, is replicated from Panel A of Figure 3.3. It illustrates the adjusted mean of the endogenous variable (y-axis) for different values of the instrument (x-axis). These adjusted means stem from a non-parametric relationship between the two variables. Additionally, vertical lines are plotted to depict the 95% confidence interval of predicted peer outcomes for different parametric specifications between the variables, as detailed in the legend. To maintain consistency, the gender composition of peers is modelled similarly to the gender composition of peers of peers (IV). The contribution from other covariates in the model is kept fixed at their average values, calculated as the mean of the product between the estimated coefficient and the observed values.

APPENDIX C: ADDITIONAL TABLES AND FIGURES (CHAPTER 4)

Table C.1: Policy effects on C-section likelihood, accounting for threshold update

	C-section likelihood					
	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline exposure measure, EXP			Exposure measure based on 37% cap		
	Baseline	Drops obs after cap revision	Models period after cap revision	Baseline (based on revised cap)	Models period after cap revision	Uses both exposure measures
$EXP * Pos$	-0.248***	-0.230***	-0.229***			-0.226***
	(0.014)	(0.013)	(0.014)			(0.014)
$EXP * Pos_{37}$			-0.044***			-0.047***
			(0.009)			(0.008)
$EXP_{37} * Pos$				-0.223***	-0.205***	
				(0.012)	(0.012)	
$EXP_{37} * Pos_{37}$					-0.043***	
					(0.008)	
Pos	-0.001	0.000	-0.000	0.002	0.002	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Pos_{37}			-0.002		-0.001	-0.001
			(0.001)		(0.001)	(0.001)
N. obs.	5,976,029	4,653,696	5,976,029	5,976,029	5,976,029	5,976,029
Municipality FE	X	X	X	X	X	X

Notes: Observations refer to individual-level (live) births that took place in the period between the 12 months before and the 12 months after the policy announcement on May 29, 1998. Births include all childbirth deliveries presented in Table 4.1. The outcome variable is the likelihood of C-section delivery. The first estimation replicates results presented in the first column of Table 4.2. Model (2) excludes observations after January 1999, when the threshold was reduced from 40% to 37%, from the estimation. Model (3) includes an indicator variable for months after January 1999 and an interaction term between this indicator variable and our treatment intensity variable, EXP. Models (4) and (5) replicate Models (1) and (3) while considering an alternative version of EXP based on the updated threshold. Model (6) incorporates both exposure measures for the corresponding periods when each threshold was in place. Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

Table C.2: Policy effects on mortality, alternative models

	Number of deaths		
	Stillbirths	Infant	Maternal
	Conditional on > 0		
$EXP * Pos$	-0.538	0.865	-0.610
	(1.452)	(1.547)	(0.960)
Pos	0.445	-1.316	0.112
	(0.721)	(0.817)	(0.258)
Baseline mean	*21.91*	*31.79*	*4.40*
	Ln(Number of deaths)		
$EXP * Pos$	-0.204*	-0.234**	-0.291
	(0.114)	(0.096)	(0.321)
Pos	0.050**	-0.032*	-0.014
	(0.022)	(0.017)	(0.056)
N. obs.	3,482	4,092	454
Municipality FE	X	X	X

Notes: Observations from death certificates are aggregated at the municipality-time level. The table presents results based on the estimation of Equation (4.3), where the outcome corresponds to the number of registered deaths of fetuses, infants, and mothers. For each outcome, we perform two regressions: one excluding observations with no reported deaths, and another with a log-transformed outcome variable. Infant mortality refers to the death of children within their first year of life, while maternal mortality relates to the death of women due to reasons associated with childbirth. Infant mortality refers to the death of children within their first year of life, while maternal mortality relates to the death of women due to reasons associated with childbirth. For stillbirths and infant deaths, time refers to childbirth deliveries within the 12-month periods before or after the policy announcement on May 29, 1998. For maternal mortality, time represents deaths occurring during the puerperium within the 12-month periods before or after the policy announcement. Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

Table C.3: Policy effects on SUS hospitalizations during first year of life, alternative models

	Number of Infant Hospitalizations (≤1yo)				
		Age at admission			
	All	0-3 months	3-6 months	6-9 months	9-12 months
	Conditional on > 0				
$EXP * Pos$	-0.538	0.865	-0.610	-0.538	0.865
	(1.452)	(1.547)	(0.960)	(1.452)	(1.547)
Pos	0.445	-1.316	0.112	0.445	-1.316
	(0.721)	(0.817)	(0.258)	(0.721)	(0.817)
Baseline mean	*21.91*	*31.79*	*4.40*	*21.91*	*31.79*
	Ln(Number of Hospitalization)				
$EXP * Pos$	-0.261***	-0.213**	-0.287***	-0.025	-0.162**
	(0.069)	(0.083)	(0.084)	(0.081)	(0.082)
Pos	0.033**	0.053***	0.009	-0.010	0.008
	(0.014)	(0.017)	(0.016)	(0.016)	(0.016)
N. obs.	5,370	4,778	4,940	5,066	5,032
Municipality FE	X	X	X	X	X

Notes: Observations from SUS hospital claims are aggregated at the municipality-time level, where time refers to indicators of patients' date of birth during the 12-month periods before or after the policy announcement on May 29, 1998. The table presents results based on the estimation of Equation (4.3), where the outcome corresponds to the number of infant hospitalizations (i.e., admissions up to 365 days old). The column titles describe the time horizon after birth for which the outcome is evaluated. For each outcome, we perform two regressions: one excluding observations with no reported hospitalization, and another with a log-transformed outcome variable. Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

Table C.4: Policy effects on share of births and infant hospitalizations in mothers' municipality of residence

	Share of births in mothers' municipality of residence			Share of infant hospitalizations in mothers' municipality of residence		
	Linear in exposure	Any exposure	By exposure quartile	Linear in exposure	Any exposure	By exposure quartile
$EXP * Pos$	0.000			0.006		
	(0.014)			(0.016)		
$1_{EXP^+} * Pos$		-0.006			-0.002	
		(0.004)			(0.005)	
$1_{EXP^+\ in\ Q1} * Pos$			-0.013**			-0.010
			(0.005)			(0.007)
$1_{EXP^+\ in\ Q2} * Pos$			-0.009			-0.001
			(0.006)			(0.006)
$1_{EXP^+\ in\ Q3} * Pos$			-0.001			-0.001
			(0.006)			(0.008)
$1_{EXP^+\ in\ Q4} * Pos$			0.000			0.004
			(0.008)			(0.009)
Pos	0.010***	0.012***	0.012***	-0.001	-0.003*	-0.003*
	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)
N. obs.	5,518	5,518	5,518	5,976,029	5,976,029	5,976,029
Municipality FE	X	X	X	X	X	X

Notes: The table presents results based on the estimation of Equation (4.3). Outcome variables are described at the top of the table. Information on the share of births and infant hospitalizations (children admitted up to 365 days old) taking place in the mothers' municipality of residence are extracted from birth certificates and SUS hospital claims, respectively. For each outcome variable, the first estimation is based on the estimation of Equation (4.3). The second estimation replaces the continuous values of the policy exposure measure with an indicator variable of positive values of exposure (i.e., EXP>0). The third estimation considers, instead, indicator variables for each quartile of the distribution of the exposure measure among municipalities where the threshold was binding. The omitted category of all estimations refers to municipalities where the introduced threshold was deemed non-binding (i.e., EXP = 0). Standard errors, presented in parentheses, are clustered at the health center level. The stars next to the estimated coefficients follow the usual convention (*** p<0.01, ** p<0.05, * p<0.1).

APPENDIX D: ADDITIONAL TABLES AND FIGURES (CHAPTER 5)

Table D.1: List of inconvenient dates between 2006-2019

	Inauspicious	Congress	Bank holiday	
			One-day	Two-day
Year 2006	13/Jan 01/Apr 13/Oct		14/Apr 21/Apr 01/May 15/Jun 07/Sep 12/Oct 02/Nov 15/Nov 25/Dec	27/Feb - 28/Feb
Year 2007	01/Apr 13/Apr 13/Jul	13/Nov - 17/Nov	01/Jan 06/Apr 01/May 07/Jun 07/Sep 12/Oct 02/Nov 15/Nov 25/Dec	19/Feb - 20/Feb
Year 2008	29/Feb 01/Apr 13/Jun		01/Jan 21/Mar 21/Apr 01/May 22/May 25/Dec	04/Feb - 05/Feb
Year 2009	13/Feb 13/Mar 01/Apr 13/Nov	14/Nov - 17/Nov	01/Jan 10/Apr 21/Apr 01/May 11/Jun 07/Sep 12/Oct 02/Nov 25/Dec	23/Feb - 24/Feb
Year 2010	01/Apr 13/Aug		01/Jan 02/Apr 21/Apr 03/Jun 07/Sep 12/Oct 02/Nov 15/Nov	15/Feb - 16/Feb
Year 2011	01/Apr 13/May	12/Nov - 15/Nov	23/Jun 07/Sep 12/Oct 02/Nov 15/Nov	07/Mar - 08/Mar 21/Apr - 22/Apr
Year 2012	13/Jan 29/Feb 01/Apr 13/Apr 13/Jul		06/Apr 01/May 07/Jun 07/Sep 12/Oct 02/Nov 15/Nov 25/Dec	20/Feb - 21/Feb

	Inauspicious	Congress	Bank holiday	
			One-day	Two-day
Year 2013	01/Apr	13/Nov - 16/Nov	01/Jan	11/Feb - 12/Feb
	13/Sep		29/Mar	
	13/Dec		01/May	
			30/May	
			15/Nov	
			25/Dec	
Year 2014	01/Apr		01/Jan	03/Mar - 04/Mar
	13/Jun		18/Apr	
			21/Apr	
			01/May	
			19/Jun	
			25/Dec	
Year 2015	13/Feb	12/Nov - 15/Nov	01/Jan	16/Feb - 17/Feb
	13/Mar		03/Apr	
	01/Apr		21/Apr	
	13/Nov		01/May	
			04/Jun	
			07/Sep	
			12/Oct	
			02/Nov	
			25/Dec	
Year 2016	29/Feb		01/Jan	08/Feb - 09/Feb
	01/Apr		25/Mar	
	13/May		21/Apr	
			26/May	
			07/Sep	
			12/Oct	
			02/Nov	
			15/Nov	
Year 2017	13/Jan	15/Nov - 18/Nov	14/Apr	27/Feb - 28/Feb
	01/Apr		21/Apr	
	13/Oct		01/May	
			15/Jun	
			07/Sep	
			12/Oct	
			02/Nov	
			15/Nov	
			25/Dec	
Year 2018	01/Apr		01/Jan	12/Feb - 13/Feb
	13/Apr		30/Mar	
	13/Jul		01/May	
			31/May	
			07/Sep	
			12/Oct	
			02/Nov	
			15/Nov	
			25/Dec	
Year 2019	01/Apr	13/Nov - 16/Nov	01/Jan	04/Mar - 05/Mar
	13/Sep		19/Apr	
	13/Dec		01/May	
			20/Jun	
			15/Nov	
			25/Dec	

Notes: This table lists all inconvenient dates considered in this study, by type of inconvenience. Dates are presented as ranges whenever they fall in subsequent days. Obstetricians-Gynaecologists National Congress lasted for 4 days in each year it took place, expect for 2007 when it lasted for 5 days. Two-day bank holidays typically refer to Carnival, which happen on the Monday and the Tuesday falling 39 and 40 days before Palm Sunday (i.e., Sunday before Easter). In 2011, it also included two separate bank holidays which happened to fall on subsequent dates. We restrict bank holidays to those falling on weekdays (i.e., Monday to Friday, which would have been business days if it was not for the official bank holiday). For instance, 1st of January of 2006 was not included above because it fell on Sunday.

Table D.2: Regression coefficients of days around *inauspicious dates:*
Number of births, delivery type, and risk profile

	Effects of Days around *Inauspicious days*											
Y	Panel A: N. deliveries				Panel B: C-section likelihood				Panel C: Risk indicator			
	Public		Private		Public		Private		Public		Private	
	White	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*
$-d_9$	0.005**	0.011***	0.006	0.000	-0.001	0.001	0.000	-0.006*	-0.002	-0.002*	-0.003*	0.000
	(0.002)	(0.004)	(0.010)	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
$-d_8$	0.006***	0.011***	-0.019**	-0.021***	-0.001	-0.004***	-0.001	0.000	-0.003**	-0.002**	-0.002*	0.002
	(0.002)	(0.004)	(0.008)	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
$-d_7$	-0.002	0.005	0.005	0.003	-0.006**	-0.002	0.005***	-0.003	-0.001	0.001	0.001	0.000
	(0.002)	(0.004)	(0.008)	(0.006)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
$-d_6$	0.001	0.006*	0.018**	0.011**	-0.004	0.002	0.000	-0.003	0.002	-0.001*	-0.001	0.004
	(0.002)	(0.004)	(0.008)	(0.005)	(0.003)	(0.002)	(0.002)	(0.004)	(0.001)	(0.001)	(0.002)	(0.003)
$-d_5$	0.002	0.006*	-0.003	-0.001	0.002	0.002	0.002	0.004	0.001	0.001	-0.001	-0.002
	(0.002)	(0.003)	(0.007)	(0.005)	(0.003)	(0.001)	(0.002)	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)
$-d_4$	0.003	0.009**	0.026***	0.024***	0.005*	0.000	0.002	0.006***	0.002	0.000	0.001	-0.003
	(0.002)	(0.004)	(0.009)	(0.006)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$-d_3$	0.007***	0.011***	0.071***	0.024***	0.006**	0.003*	0.005***	0.001	0.000	0.001	-0.003**	-0.002
	(0.002)	(0.004)	(0.011)	(0.006)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
$-d_2$	0.001	0.007**	0.004	0.005	0.001	-0.001	-0.003	0.001	0.001	-0.001	-0.001	-0.001
	(0.002)	(0.004)	(0.008)	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.003)
$-d_1$	0.002	0.001	0.058***	-0.002	0.002	-0.001	0.005***	0.000	0.002*	0.000	0.000	-0.005**
	(0.002)	(0.004)	(0.010)	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
$\boldsymbol{d_1}$	**-0.003**	**-0.001**	**-0.269*** **	**-0.066*** **	**-0.002**	**-0.003** **	**-0.022*** **	**-0.014*** **	**0.000**	**0.001**	**0.002**	**-0.001**
	(0.002)	**(0.004)**	**(0.024)**	**(0.008)**	**(0.003)**	**(0.002)**	**(0.002)**	**(0.003)**	**(0.001)**	**(0.001)**	**(0.002)**	**(0.002)**
$+d_1$	-0.002	-0.005	0.067***	0.029***	0.001	0.001	0.007***	0.008**	0.003*	0.000	0.000	-0.001
	(0.002)	(0.004)	(0.010)	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
$+d_2$	-0.001	0.000	0.015*	0.008*	0.003	0.002	0.003	0.003	0.000	-0.000	-0.000	-0.001
	(0.002)	(0.004)	(0.008)	(0.005)	(0.003)	(0.002)	(0.003)	(0.004)	(0.001)	(0.001)	(0.002)	(0.003)
$+d_3$	0.002	0.010***	0.064***	0.024***	0.003	0.001	0.002	-0.000	-0.000	0.002**	0.000	0.001
	(0.002)	(0.004)	(0.009)	(0.006)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
$+d_4$	0.001	0.005	0.006	-0.012**	-0.001	-0.001	0.001	-0.002	0.000	0.001	0.000	-0.001
	(0.002)	(0.004)	(0.008)	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
$+d_5$	0.003	-0.005	0.028***	0.005	-0.001	-0.002*	-0.000	-0.003	0.001	0.000	-0.004**	0.003
	(0.002)	(0.004)	(0.009)	(0.005)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
$+d_6$	0.001	0.002	-0.006	-0.015***	-0.003	0.002	-0.001	0.002	-0.000	-0.001	-0.003*	-0.003
	(0.002)	(0.004)	(0.009)	(0.006)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)

Effects of Days around *Inauspicious days*												
Y	Panel A: N. deliveries				Panel B: C-section likelihood				Panel C: Risk indicator			
	Public		Private		Public		Private		Public		Private	
	White	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*
$+d_7$	0.001	0.014***	0.086***	0.033***	0.002	0.005***	0.004***	0.002	-0.001	0.001	-0.001	-0.003
	(0.002)	(0.004)	(0.011)	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
$+d_8$	-0.000	0.005	0.009	0.009*	-0.002	0.001	0.003	-0.004	-0.000	-0.000	-0.002	-0.000
	(0.002)	(0.003)	(0.008)	(0.005)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
$+d_9$	0.000	0.003	0.008	0.015***	0.000	0.003*	0.005*	-0.004	0.000	0.000	0.002	0.002
	(0.002)	(0.004)	(0.006)	(0.005)	(0.003)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.002)	(0.003)
N.obs	11,937,950	11,937,950	2,998,166	2,998,166	3,946,411	12,107,865	3,903,499	1,748,534	3,950,171	12,122,699	3,905,811	1,750,163

Notes: Each column reports point estimates and standard errors (in parenthesis) of a regression that considers the sample of mothers of a given race (white, black) in a given type of hospital (public, private). Panel titles indicate outcome variables. Unit level is individual birth for regression results reported in Panels B and C and hospital-day-race for those reported in Panel A. Risk indicator is classified as multiple pregnancy, non-cephalic foetus presentation or newborn with congenital anomaly. Standard errors are clustered at the hospital level. For more details on regression specification, see Equation (5.1) in Section 5.2. The coefficient term -d_n ($+d_n$) refers to the n[th] day before the beginning (after the end) of the inconvenient day(s) under consideration. This table shows coefficient estimates around the *inauspicious day*, represented by d_1.

Table D.3: Regression coefficients of days around *Obstetricians-Gynaecologists:* Number of births, delivery type, and risk profile

Effects of Days around *Obstetricians-Gynaecologists Congress*												
Y	Panel A: N. deliveries				Panel B: C-section likelihood				Panel C: Risk indicator			
	Public		Private		Public		Private		Public		Private	
	White	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*
-d_9	0.003	0.021**	0.145***	0.124***	0.017**	0.009**	0.005	-0.004	0.003	0.002	0.000	0.008
	(0.005)	(0.010)	(0.024)	(0.018)	(0.008)	(0.004)	(0.004)	(0.006)	(0.004)	(0.002)	(0.004)	(0.006)
-d_8	0.007	0.035***	0.156***	0.060***	0.007	0.018***	0.006	0.006	-0.002	0.001	0.000	-0.010**
	(0.005)	(0.009)	(0.026)	(0.015)	(0.007)	(0.004)	(0.004)	(0.006)	(0.003)	(0.002)	(0.004)	(0.005)
-d_7	-0.001	0.021**	0.106***	0.091***	0.016**	0.005	0.003	0.014**	0.002	-0.001	-0.001	-0.011**
	(0.005)	(0.009)	(0.023)	(0.016)	(0.007)	(0.004)	(0.005)	(0.006)	(0.004)	(0.002)	(0.004)	(0.005)
-d_6	0.010**	0.037***	0.076***	0.093***	0.008	0.011***	0.006	0.009	0.001	0.002	-0.001	-0.001
	(0.005)	(0.010)	(0.020)	(0.016)	(0.007)	(0.004)	(0.005)	(0.006)	(0.004)	(0.002)	(0.004)	(0.005)
-d_5	0.009**	0.023**	0.170***	0.022	0.014**	0.012***	0.005	0.008	0.001	0.005**	0.005	-0.010*
	(0.005)	(0.010)	(0.026)	(0.015)	(0.007)	(0.004)	(0.004)	(0.007)	(0.004)	(0.002)	(0.004)	(0.005)
-d_4	0.001	0.015*	0.172***	0.025*	0.007	0.014***	0.003	0.003	0.000	0.000	0.000	-0.002
	(0.005)	(0.009)	(0.026)	(0.014)	(0.007)	(0.004)	(0.004)	(0.008)	(0.004)	(0.002)	(0.004)	(0.006)
-d_3	0.009*	0.013	0.148***	0.063***	0.002	0.012***	0.009*	0.011	-0.002	0.006***	-0.002	-0.008
	(0.005)	(0.009)	(0.027)	(0.015)	(0.007)	(0.004)	(0.005)	(0.007)	(0.003)	(0.002)	(0.004)	(0.005)
-d_2	0.006	0.040***	0.157***	0.110***	0.016**	0.010**	0.006	0.008	0.003	-0.001	-0.001	0.001
	(0.005)	(0.010)	(0.028)	(0.018)	(0.007)	(0.004)	(0.004)	(0.006)	(0.004)	(0.002)	(0.004)	(0.005)

Effects of Days around *Obstetricians-Gynaecologists Congress*												
Y	Panel A: N. deliveries				Panel B: C-section likelihood				Panel C: Risk indicator			
	Public		Private		Public		Private		Public		Private	
	White	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*
$-d_1$	0.009*	0.031***	0.136***	0.071***	0.002	0.007*	0.009**	0.014**	0.003	0.001	0.001	0.003
	(0.005)	(0.009)	(0.023)	(0.016)	(0.007)	(0.004)	(0.005)	(0.006)	(0.004)	(0.002)	(0.004)	(0.006)
$\boldsymbol{d_1}$	-0.001	-0.009	-0.092***	-0.025**	0.004	0.003	-0.011**	0.004	0.003	-0.001	-0.006	-0.005
	(0.005)	(0.009)	(0.020)	(0.013)	(0.007)	(0.004)	(0.005)	(0.008)	(0.004)	(0.002)	(0.004)	(0.006)
$\boldsymbol{d_2}$	-0.002	0.024***	-0.067***	0.036**	0.007	0.004	-0.000	-0.009	-0.005	0.003	-0.005	0.008
	(0.005)	(0.009)	(0.022)	(0.015)	(0.007)	(0.004)	(0.006)	(0.008)	(0.004)	(0.002)	(0.005)	(0.007)
$\boldsymbol{d_3}$	-0.013**	-0.026***	-0.275***	-0.109***	-0.009	-0.008*	-0.021***	-0.012	-0.000	-0.001	0.005	0.005
	(0.005)	(0.009)	(0.026)	(0.014)	(0.007)	(0.004)	(0.006)	(0.008)	(0.004)	(0.002)	(0.004)	(0.006)
$\boldsymbol{d_4}$	-0.010**	-0.035***	-0.206***	-0.096***	-0.014*	-0.001	-0.016**	-0.012	-0.000	-0.002	0.002	-0.002
	(0.005)	(0.009)	(0.022)	(0.013)	(0.008)	(0.004)	(0.007)	(0.010)	(0.004)	(0.002)	(0.005)	(0.007)
$\boldsymbol{d_5}$	-0.019	-0.024`	-0.211***	0.069***	-0.007	-0.010	-0.068***	0.036	-0.002	0.004	0.007	0.005
	(0.014)	(0.020)	(0.039)	(0.019)	(0.018)	(0.011)	(0.025)	(0.034)	(0.006)	(0.005)	(0.016)	(0.034)
$+d_1$	0.002	0.004	0.026	0.040***	-0.002	0.008**	-0.002	0.001	-0.005	-0.000	0.002	0.004
	(0.005)	(0.009)	(0.019)	(0.014)	(0.008)	(0.004)	(0.007)	(0.010)	(0.004)	(0.002)	(0.005)	(0.007)
$+d_2$	0.007	0.020**	0.025	0.066***	0.006	0.006	0.007**	0.013**	0.001	0.004**	0.007	0.002
	(0.005)	(0.009)	(0.020)	(0.016)	(0.007)	(0.004)	(0.004)	(0.006)	(0.004)	(0.002)	(0.005)	(0.005)
$+d_3$	0.002	0.021**	0.039*	0.073***	0.012*	0.011***	0.004	0.008	0.006	0.004*	0.009**	0.002
	(0.005)	(0.009)	(0.021)	(0.017)	(0.007)	(0.004)	(0.004)	(0.007)	(0.004)	(0.002)	(0.004)	(0.006)
$+d_4$	0.007	0.013	0.002	0.067***	0.002	0.002	0.001	0.015**	-0.002	-0.002	-0.003	0.005
	(0.005)	(0.009)	(0.019)	(0.016)	(0.007)	(0.004)	(0.005)	(0.006)	(0.003)	(0.002)	(0.004)	(0.006)
$+d_5$	0.008	0.028***	0.025	0.043***	0.012	0.005	0.002	-0.007	0.001	0.003*	0.002	-0.005
	(0.005)	(0.009)	(0.019)	(0.016)	(0.007)	(0.004)	(0.005)	(0.008)	(0.004)	(0.002)	(0.005)	(0.006)
$+d_6$	0.005	0.005	0.024	0.015	0.000	0.007*	-0.004	-0.009	0.006	0.000	-0.000	-0.001
	(0.005)	(0.009)	(0.020)	(0.014)	(0.007)	(0.004)	(0.005)	(0.007)	(0.004)	(0.002)	(0.004)	(0.006)
$+d_7$	0.004	0.000	0.018	0.023*	-0.002	-0.004	0.002	-0.009	-0.001	0.000	0.010**	-0.009
	(0.005)	(0.009)	(0.021)	(0.014)	(0.007)	(0.004)	(0.006)	(0.008)	(0.003)	(0.002)	(0.005)	(0.007)
$+d_8$	-0.007	0.009	0.070***	0.026**	0.010	0.008**	0.012**	0.005	0.003	-0.002	0.003	0.008
	(0.005)	(0.009)	(0.019)	(0.012)	(0.007)	(0.004)	(0.005)	(0.008)	(0.004)	(0.002)	(0.005)	(0.006)
$+d_9$	0.001	0.005	0.022	0.048***	0.004	0.011***	0.008**	-0.001	0.003	0.002	0.008**	-0.005
	(0.005)	(0.009)	(0.022)	(0.014)	(0.007)	(0.004)	(0.004)	(0.006)	(0.004)	(0.002)	(0.004)	(0.005)
N.obs	11,937,950	11,937,950	2,998,166	2,998,166	3,946,411	12,107,865	3,903,499	1,748,534	3,950,171	12,122,699	3,905,811	1,750,163

Notes: General table description can be found in notes of Table D.2. This table shows coefficient estimates around the period encompassing *congress days*, represented by d_1, d_2, d_3, d_4, d_5.

Table D.4: Regression coefficients of days around *one-day bank holidays:*
Number of births, delivery type, and risk profile

	Effects of Days around *One-day bank holidays*											
Y	Panel A: N. deliveries				Panel B: C-section likelihood				Panel C: Risk indicator			
	Public		Private		Public		Private		Public		Private	
	White	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*
$-d_9$	-0.002	0.003	-0.027***	0.004	-0.001	-0.001	-0.002**	-0.002	0.001	0.000	0.002*	-0.003***
	(0.001)	(0.002)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$-d_8$	-0.004***	-0.002	-0.059***	-0.014***	0.001	-0.002*	-0.003**	-0.003*	-0.000	-0.000	0.000	-0.004**
	(0.001)	(0.002)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$-d_7$	-0.003**	-0.002	-0.023***	-0.013***	-0.000	-0.002*	0.000	-0.001	-0.000	-0.001	0.001	0.000
	(0.001)	(0.002)	(0.005)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)
$-d_6$	0.003**	0.002	0.120***	0.044***	0.003*	0.003***	0.009***	0.006***	0.000	0.001**	-0.001	-0.001
	(0.001)	(0.003)	(0.012)	(0.005)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$-d_5$	0.002	0.002	0.128***	0.040***	0.002	0.000	0.009***	0.009***	0.001	0.000	-0.003***	-0.002
	(0.001)	(0.002)	(0.013)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$-d_4$	-0.000	0.004*	0.109***	0.037***	0.002	0.002	0.007***	0.006***	0.001	-0.001***	-0.000	0.001
	(0.001)	(0.002)	(0.010)	(0.005)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$-d_3$	0.003**	0.007***	0.092***	0.032***	0.004**	0.002**	0.007***	0.009***	0.000	0.000	0.000	0.002
	(0.001)	(0.002)	(0.008)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$-d_2$	0.000	0.001	-0.030***	0.004	0.003**	0.001	-0.004***	-0.002	-0.001	-0.001	0.001	0.000
	(0.001)	(0.002)	(0.009)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$-d_1$	-0.008***	-0.013***	-0.148***	-0.056***	-0.002	-0.006***	-0.010***	-0.009***	-0.002*	-0.001*	0.000	0.002
	(0.001)	(0.003)	(0.017)	(0.007)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$\boldsymbol{d_1}$	-0.047***	-0.097***	-0.622***	-0.253***	-0.054***	-0.044***	-0.049***	-0.044***	-0.005***	-0.004***	0.000	0.000
	(0.003)	(0.004)	(0.043)	(0.014)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
$+d_1$	-0.011***	-0.018***	-0.067***	-0.021***	-0.015***	-0.016***	-0.002	-0.005***	-0.001	-0.001*	-0.000	0.002
	(0.002)	(0.003)	(0.011)	(0.005)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$+d_2$	0.000	0.006**	-0.009	0.007**	-0.002	-0.000	-0.000	0.002	0.001	0.001	0.002	-0.003*
	(0.001)	(0.003)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)
$+d_3$	0.005***	0.014***	0.047***	0.027***	0.001	0.002**	0.005***	0.000	0.001	0.000	-0.001	0.000
	(0.001)	(0.002)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$+d_4$	0.004***	0.008***	0.025***	0.011***	0.002	0.002**	0.001	0.000	-0.000	0.001*	0.000	-0.000
	(0.001)	(0.002)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)	(0.001)
$+d_5$	0.003**	0.009***	-0.006	0.012***	-0.001	0.001	-0.002	-0.000	-0.001	-0.001	0.002**	-0.003**
	(0.001)	(0.002)	(0.005)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$+d_6$	-0.000	-0.002	-0.068***	-0.015***	-0.003*	-0.002	-0.004***	-0.005***	0.002*	-0.001	-0.001	-0.002
	(0.001)	(0.002)	(0.009)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)

Y	Panel A: N. deliveries				Panel B: C-section likelihood				Panel C: Risk indicator			
	Public		Private		Public		Private		Public		Private	
	White	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*
	Effects of Days around *One-day bank holidays*											
$+d_7$	-0.005***	-0.006***	-0.063***	-0.012***	-0.002	-0.000	-0.002*	-0.006***	-0.001	0.000	-0.000	-0.000
	(0.001)	(0.002)	(0.007)	(0.003)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$+d_8$	-0.001	0.003	0.000	0.007*	0.003*	0.001	0.002*	0.002	-0.000	-0.001	-0.001	-0.001
	(0.001)	(0.002)	(0.005)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$+d_9$	0.002*	0.006**	0.032***	0.016***	-0.002	0.002	0.005***	0.004*	-0.001	0.001	-0.001	-0.003*
	(0.001)	(0.002)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
N.obs	11,937,950	11,937,950	2,998,166	2,998,166	3,946,411	12,107,865	3,903,499	1,748,534	3,950,171	12,122,699	3,905,811	1,750,163

Notes: General table description can be found in notes of Table D.2. This table shows coefficient estimates around the *bank holiday*, represented by d_1.

Table D.5: Regression coefficients of days around *two-day bank holidays:* Number of births, delivery type, and risk profile

Y	Panel A: N. deliveries				Panel B: C-section likelihood				Panel C: Risk indicator			
	Public		Private		Public		Private		Public		Private	
	White	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*
	Effects of Days around *Two-day bank holidays*											
$-d_9$	-0.001	0.003	0.062***	0.013	0.011**	0.004	0.005	0.004	0.004*	-0.001	-0.006*	-0.002
	(0.003)	(0.005)	(0.017)	(0.008)	(0.005)	(0.003)	(0.004)	(0.006)	(0.002)	(0.001)	(0.003)	(0.004)
$-d_8$	0.001	-0.001	-0.012	-0.003	-0.006	-0.002	0.002	-0.008	0.001	0.001	-0.006*	-0.004
	(0.003)	(0.005)	(0.011)	(0.008)	(0.005)	(0.003)	(0.005)	(0.008)	(0.002)	(0.001)	(0.003)	(0.005)
$-d_7$	0.008**	-0.010	0.129***	0.038***	0.000	0.001	0.007***	0.008*	0.000	-0.003**	-0.001	-0.006
	(0.004)	(0.006)	(0.019)	(0.011)	(0.004)	(0.002)	(0.002)	(0.004)	(0.002)	(0.001)	(0.002)	(0.003)
$-d_6$	0.010**	-0.006	0.204***	0.058***	0.001	0.004	0.009***	0.012***	0.001	0.002	-0.002	-0.007*
	(0.004)	(0.006)	(0.021)	(0.010)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)	(0.001)	(0.002)	(0.004)
$-d_5$	0.010***	-0.001	0.281***	0.086***	0.003	0.004	0.014***	0.007	0.000	0.001	-0.003	0.002
	(0.003)	(0.006)	(0.025)	(0.011)	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)	(0.001)	(0.003)	(0.003)
$-d_4$	0.010***	-0.001	0.250***	0.086***	0.008*	0.011***	0.011***	0.020***	-0.005**	0.001	-0.001	-0.000
	(0.004)	(0.006)	(0.023)	(0.012)	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)	(0.001)	(0.002)	(0.004)
$-d_3$	0.011***	0.013**	0.102***	0.033***	0.010**	0.007**	-0.002	0.009*	-0.000	0.000	-0.002	0.004
	(0.004)	(0.006)	(0.014)	(0.012)	(0.004)	(0.003)	(0.003)	(0.005)	(0.002)	(0.001)	(0.002)	(0.004)
$-d_2$	-0.002	-0.005	-0.078***	-0.050***	-0.004	0.001	-0.013***	-0.013*	0.003	-0.002	-0.001	0.009*
	(0.003)	(0.006)	(0.015)	(0.010)	(0.005)	(0.003)	(0.004)	(0.007)	(0.002)	(0.001)	(0.003)	(0.006)
$-d_1$	0.010***	0.010	-0.138***	-0.067***	0.009*	0.009***	-0.014***	-0.011	0.002	0.001	0.003	-0.007
	(0.004)	(0.006)	(0.020)	(0.010)	(0.005)	(0.003)	(0.005)	(0.010)	(0.003)	(0.001)	(0.004)	(0.005)

Y	Panel A: N. deliveries				Panel B: C-section likelihood				Panel C: Risk indicator			
	Public		Private		Public		Private		Public		Private	
	White	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*	*White*	*Black*
Effects of Days around *Two-day bank holidays*												
$\boldsymbol{d_1}$	-0.028***	-0.072***	-0.677***	-0.321***	-0.039***	-0.036***	-0.041***	-0.038***	-0.005*	-0.003***	0.008**	-0.003
	(0.004)	(0.007)	(0.051)	(0.025)	(0.005)	(0.003)	(0.004)	(0.008)	(0.002)	(0.001)	(0.003)	(0.005)
$\boldsymbol{d_2}$	-0.053***	-0.121***	-0.790***	-0.350***	-0.064***	-0.053***	-0.069***	-0.072***	-0.004*	-0.005***	0.008**	0.003
	(0.004)	(0.008)	(0.059)	(0.023)	(0.006)	(0.003)	(0.005)	(0.009)	(0.002)	(0.001)	(0.004)	(0.006)
$+d_1$	-0.022***	-0.078***	-0.143***	-0.087***	-0.036***	-0.039***	-0.003	-0.017***	-0.002	-0.003*	0.001	0.004
	(0.004)	(0.007)	(0.020)	(0.013)	(0.005)	(0.003)	(0.003)	(0.005)	(0.002)	(0.001)	(0.003)	(0.004)
$+d_2$	0.011***	0.012*	0.151***	0.080***	0.001	-0.003	0.005	0.005	0.001	0.000	-0.002	0.002
	(0.004)	(0.006)	(0.018)	(0.012)	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)	(0.001)	(0.002)	(0.004)
$+d_3$	0.017***	0.048***	0.172***	0.065***	0.005	0.007**	0.003	0.008*	0.006***	-0.001	-0.004*	0.002
	(0.004)	(0.006)	(0.019)	(0.011)	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)	(0.001)	(0.002)	(0.004)
$+d_4$	0.015***	0.024***	0.090***	0.037***	0.004	0.006**	0.008**	0.004	-0.005***	0.001	0.003	-0.001
	(0.004)	(0.006)	(0.015)	(0.009)	(0.005)	(0.003)	(0.004)	(0.005)	(0.002)	(0.001)	(0.003)	(0.004)
$+d_5$	0.006**	0.013**	0.017	0.007	0.007	0.007**	0.003	0.006	0.001	0.000	0.000	0.005
	(0.003)	(0.006)	(0.011)	(0.007)	(0.005)	(0.003)	(0.004)	(0.007)	(0.002)	(0.001)	(0.003)	(0.005)
$+d_6$	0.011***	0.014**	0.025	0.015	-0.000	-0.001	0.006**	0.000	-0.002	-0.000	-0.004	-0.006*
	(0.003)	(0.006)	(0.015)	(0.010)	(0.004)	(0.002)	(0.002)	(0.004)	(0.002)	(0.001)	(0.002)	(0.003)
$+d_7$	0.009**	-0.000	0.000	0.008	-0.004	0.002	-0.003	-0.002	-0.000	0.000	-0.001	-0.000
	(0.004)	(0.006)	(0.014)	(0.009)	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)	(0.001)	(0.002)	(0.004)
$+d_8$	0.009**	0.006	-0.074***	-0.003	-0.002	0.002	-0.008***	0.005	0.000	-0.003**	-0.001	0.000
	(0.004)	(0.006)	(0.015)	(0.010)	(0.004)	(0.002)	(0.003)	(0.004)	(0.002)	(0.001)	(0.002)	(0.004)
$+d_9$	0.014***	0.005	0.065***	0.023**	0.010**	-0.000	0.001	-0.001	0.000	0.001	-0.004*	-0.010***
	(0.004)	(0.006)	(0.014)	(0.010)	(0.004)	(0.003)	(0.003)	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)
N.obs	11,937,950	11,937,950	2,998,166	2,998,166	3,946,411	12,107,865	3,903,499	1,748,534	3,950,171	12,122,699	3,905,811	1,750,163

Notes: General table description can be found in notes of Table D.2. This table shows coefficient estimates around the period encompassing *two-day bank holidays*, represented by d_1, d_2.

Table D.6: Regression coefficients of days around inconvenient periods: Excess births of back mothers

Y	Difference in Number of Black vs White Births							
	Inauspicious days		Congress days		One-day bank holidays		Two-day bank holidays	
	Public	*Private*	*Public*	*Private*	*Public*	*Private*	*Public*	*Private*
$-d_9$	0.006	-0.006	0.018*	-0.022	0.004	0.031***	0.005	-0.048***
	(0.004)	(0.011)	(0.011)	(0.029)	(0.003)	(0.007)	(0.006)	(0.019)
$-d_8$	0.005	-0.002	0.028***	-0.096***	0.002	0.046***	-0.002	0.009
	(0.004)	(0.010)	(0.010)	(0.027)	(0.003)	(0.007)	(0.006)	(0.013)
$-d_7$	0.007	-0.002	0.022**	-0.015	0.001	0.010	-0.018**	-0.091***
	(0.005)	(0.010)	(0.010)	(0.026)	(0.003)	(0.006)	(0.007)	(0.021)
$-d_6$	0.005	-0.006	0.026**	0.017	-0.000	-0.076***	-0.016**	-0.146***
	(0.004)	(0.009)	(0.011)	(0.025)	(0.003)	(0.011)	(0.007)	(0.022)
$-d_5$	0.004	0.002	0.014	-0.149***	-0.000	-0.089***	-0.011*	-0.195***
	(0.004)	(0.009)	(0.011)	(0.030)	(0.003)	(0.011)	(0.007)	(0.024)
$-d_4$	0.006	-0.002	0.014	-0.147***	0.005*	-0.071***	-0.011	-0.163***
	(0.004)	(0.011)	(0.011)	(0.029)	(0.003)	(0.010)	(0.007)	(0.023)
$-d_3$	0.004	-0.047***	0.004	-0.085***	0.004	-0.061***	0.003	-0.069***
	(0.004)	(0.012)	(0.010)	(0.029)	(0.003)	(0.008)	(0.007)	(0.018)
$-d_2$	0.006	0.001	0.034***	-0.047	0.001	0.034***	-0.003	0.028*
	(0.004)	(0.010)	(0.011)	(0.032)	(0.003)	(0.009)	(0.007)	(0.015)
$-d_1$	-0.002	-0.060***	0.022**	-0.065**	-0.006*	0.092***	-0.000	0.071***
	(0.004)	(0.011)	(0.011)	(0.027)	(0.003)	(0.016)	(0.007)	(0.017)
$\boldsymbol{d_1}$	0.002	0.203***	-0.007	0.067***	-0.050***	0.369***	-0.044***	0.357***
	(0.004)	(0.022)	(0.010)	(0.023)	(0.004)	(0.040)	(0.008)	(0.046)
$\boldsymbol{d_2}$			0.027**	0.103***			-0.068***	0.440***
			(0.011)	(0.026)			(0.008)	(0.054)
$\boldsymbol{d_3}$			-0.013	0.166***				
			(0.010)	(0.027)				
$\boldsymbol{d_4}$			-0.025**	0.110***				
			(0.010)	(0.022)				
$\boldsymbol{d_5}$			-0.005	0.280***				
			(0.025)	(0.043)				
$+d_1$	-0.003	-0.038***	0.002	0.013	-0.007**	0.047***	-0.055***	0.056***
	(0.004)	(0.010)	(0.011)	(0.022)	(0.003)	(0.011)	(0.008)	(0.020)
$+d_2$	0.002	-0.007	0.013	0.041	0.006**	0.016**	0.000	-0.071***
	(0.004)	(0.009)	(0.011)	(0.025)	(0.003)	(0.007)	(0.007)	(0.019)
$+d_3$	0.007*	-0.040***	0.019*	0.034	0.009***	-0.020***	0.031***	-0.107***
	(0.004)	(0.011)	(0.010)	(0.026)	(0.003)	(0.007)	(0.007)	(0.022)
$+d_4$	0.004	-0.018*	0.005	0.065***	0.004	-0.014**	0.009	-0.053***
	(0.004)	(0.010)	(0.010)	(0.024)	(0.003)	(0.007)	(0.007)	(0.017)
$+d_5$	-0.008*	-0.023**	0.021**	0.018	0.007**	0.017***	0.006	-0.010
	(0.004)	(0.010)	(0.010)	(0.024)	(0.003)	(0.006)	(0.007)	(0.013)
$+d_6$	0.000	-0.009	-0.000	-0.009	-0.002	0.053***	0.003	-0.010
	(0.004)	(0.011)	(0.010)	(0.023)	(0.003)	(0.009)	(0.007)	(0.019)
$+d_7$	0.013***	-0.053***	-0.004	0.005	-0.001	0.051***	-0.009	0.008
	(0.004)	(0.012)	(0.010)	(0.023)	(0.003)	(0.008)	(0.007)	(0.017)
$+d_8$	0.005	0.000	0.016	-0.044**	0.004	0.007	-0.003	0.070***
	(0.004)	(0.010)	(0.010)	(0.021)	(0.003)	(0.007)	(0.007)	(0.018)
$+d_9$	0.002	0.007	0.004	0.027	0.004	-0.016**	-0.008	-0.043**
	(0.004)	(0.008)	(0.010)	(0.026)	(0.003)	(0.007)	(0.007)	(0.017)
N.obs	11,937,950	2,998,166	11,937,950	2,998,166	11,937,950	2,998,166	11,937,950	2,998,166

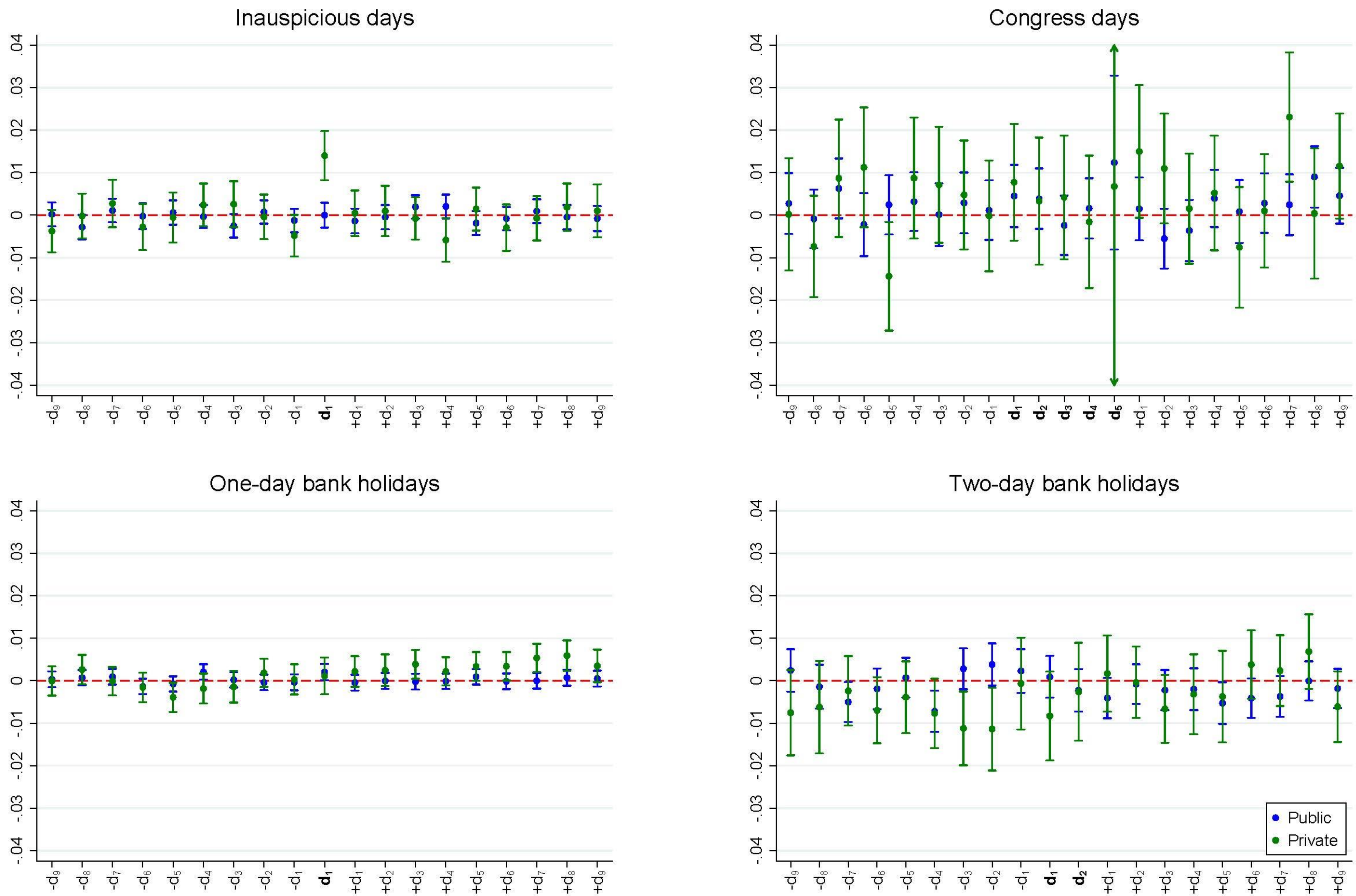
Notes: Each column reports point estimates and standard errors (in parenthesis) of a regression that considers the sample of deliveries in a given type of hospital (public, private) around a given type of inconvenient period. Outcome is the difference between the number of births delivered by black mothers and the number of births delivered by white mothers in a given hospital-day. Standard errors are clustered at the hospital level. For more details on regression specification, see Equation (5.1) in Section 5.2. The coefficient term $-d_n$ ($+d_n$) refers to the n^{th} day before the beginning (after the end) of the inconvenient day(s) under consideration.

Table D.7: Regression coefficients of inconvenient periods and SUS affiliation: Excess births of black mothers

Y	Difference in Number of Black vs White Births			
	Inauspicious days	Congress days	One-day bank holiday	Two-day bank holiday
Omitted category: D(Incon) * D(SUS==0%)				
D(Incon) * D(SUS>0% & <25%)	-0.042	-0.042	0.113	0.054
	(0.042)	(0.049)	(0.100)	(0.083)
D(Incon) *D(SUS>=25% & <50%)	-0.101***	-0.097**	-0.102**	-0.134***
	(0.028)	(0.038)	(0.043)	(0.041)
D(Incon) * D(SUS>=50% & <75%)	-0.091***	-0.044*	-0.092***	-0.138***
	(0.017)	(0.025)	(0.030)	(0.030)
D(Incon) * D(SUS>=75% & <100%)	-0.116***	-0.166***	-0.233***	-0.198***
	(0.020)	(0.027)	(0.030)	(0.033)
D(Incon) * D(SUS==100%)	-0.111***	-0.218***	-0.267***	-0.194***
	(0.015)	(0.020)	(0.026)	(0.027)
D(Incon)	0.134***	0.171***	0.245***	0.198***
	(0.015)	(0.018)	(0.027)	(0.029)
Omitted category: D(Around) * D(SUS==0%)				
D(Around) * D(SUS>0% & <25%)	-0.009	-0.063*	-0.015	-0.028
	(0.011)	(0.037)	(0.010)	(0.028)
D(Around) *D(SUS>=25% & <50%)	-0.010	-0.001	0.000	0.004
	(0.009)	(0.025)	(0.007)	(0.015)
D(Around) * D(SUS>=50% & <75%)	-0.001	0.027**	-0.001	-0.005
	(0.006)	(0.012)	(0.004)	(0.011)
D(Around) * D(SUS>=75% & <100%)	0.013*	-0.012	0.006	0.026**
	(0.007)	(0.014)	(0.005)	(0.012)
D(Around) * D(SUS==100%)	0.031***	-0.042***	-0.001	0.065***
	(0.004)	(0.009)	(0.003)	(0.009)
D(Around)	-0.022***	0.025***	0.001	-0.054***
	(0.004)	(0.008)	(0.003)	(0.008)
Omitted category: D(SUS==0%)				
D(SUS>0% & <25%)	0.155**	0.155**	0.156**	0.155**
	(0.074)	(0.074)	(0.074)	(0.074)
D(SUS>=25% & <50%)	0.274***	0.273***	0.274***	0.273***
	(0.066)	(0.066)	(0.066)	(0.066)
D(SUS>=50% & <75%)	0.301***	0.300***	0.303***	0.302***
	(0.051)	(0.051)	(0.051)	(0.051)
D(SUS>=75% & <100%)	0.293***	0.295***	0.297***	0.294***
	(0.060)	(0.060)	(0.060)	(0.060)
D(SUS==100%)	0.263***	0.269***	0.273***	0.264***
	(0.058)	(0.058)	(0.058)	(0.058)
N.obs.	23,505,846	23,505,846	23,505,846	23,505,846

Notes: The table presents point estimates and standard errors (in parenthesis) of all coefficients of Equation (5.2), which absorbs fixed effects of hospital, year, month, and weekday. Outcome is the difference between the number of births delivered by black mothers and the number of births delivered by white mothers in a given hospital-day. Each column considers a given type of inconvenient period. Unit is Standard errors are clustered at the hospital level.

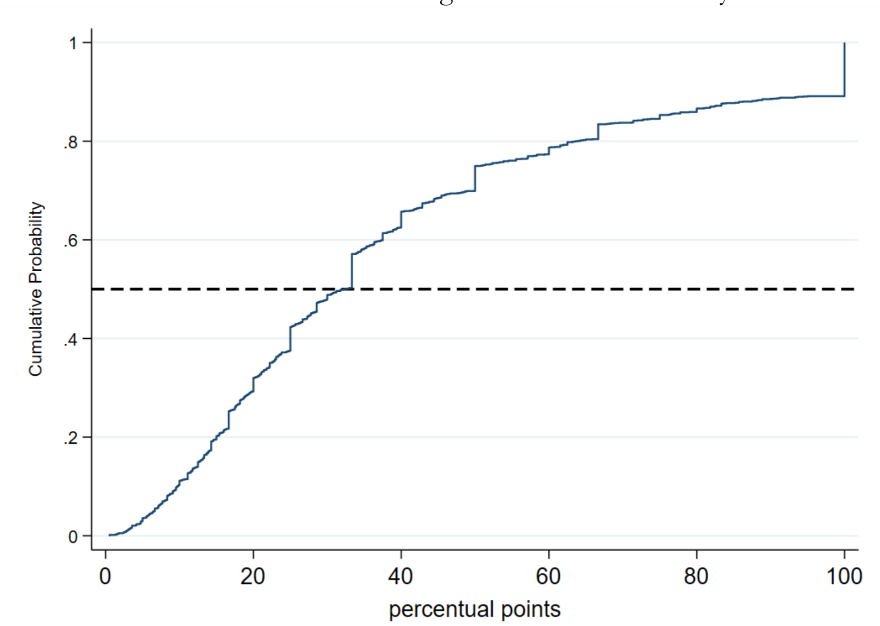
Figure D.1: Regression coefficients of days around inconvenient periods: Share of births by black mothers



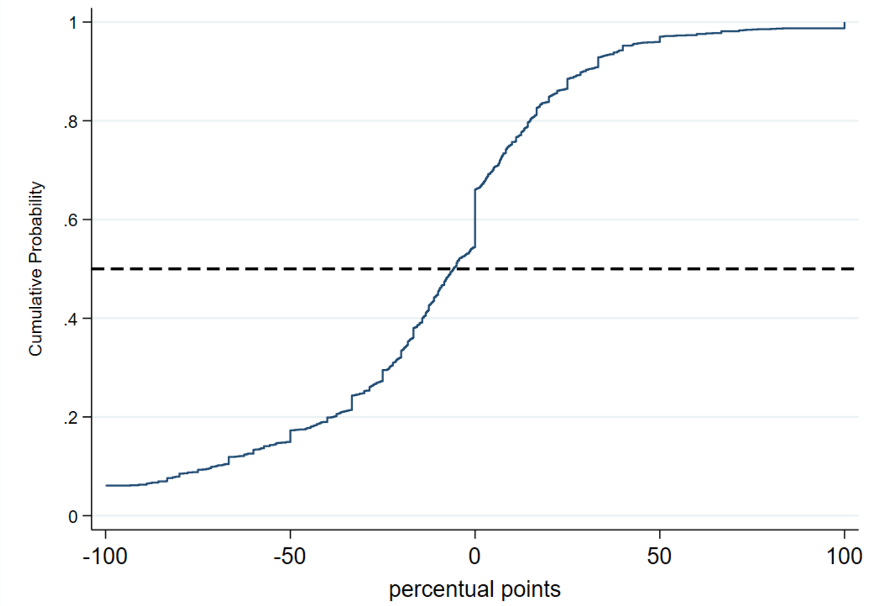
Notes: Outcome variable is the share of number of deliveries by black mothers. Unit level is hospital-day. Each plot shows point estimates and respective 95% confidence intervals for two separate regressions: one for births taking place in the Public sector (in blue) and another for those which happened in the Private sector (in green). Plot titles indicate the type of inconvenient day under consideration. Standard errors are clustered at the hospital level. For more details on regression specification, see Equation (5.1) in Section 5.2. The coefficient term -d_n (+d_n) refers to the n[th] day before the beginning (after the end) of the inconvenient period (displayed in bold).

Figure D.2: Cumulative distribution of hospital's monthly share of obstetric beds attached to SUS

Panel A: Difference between highest and lowest monthly share



Panel B: Difference between last and first reported monthly shares



Notes: Plots consider 1,654 hospitals who report at least two different monthly shares of hospital's obstetric beds attached to SUS between January 2006 and December 2019. The top plot presents the cumulative distribution of hospitals' difference between highest and lowest reported monthly shares. It ranges from just above 0 (i.e., sample comprises of hospitals presenting some variation) and 100 percentual points (i.e., hospitals who report months with no obstetric beds (0%), associated to SUS and other months with all available obstetric beds (100%) reserved to SUS). The bottom plot shows cumulative distribution of hospitals' difference between last and first reported monthly shares. It ranges from -100 to +100 percentual points, where -100 p.p. refers to facilities moving from solely serving the public sector to solely serving the private sector while +100 p.p. refers to fully private facilities in the start of our time sample who turned into fully public ones as of the last period with available information. The dashed horizontal line marks 0.5 in cumulative probability (y-axis). The point at which it crosses the cdf refers to the median value among the sample of hospitals under consideration.

8 REFERENCES

Abaluck, J., Agha, L., Chan, D., Singer, D., & Zhu, Y. (2021). Fixing Misallocation with Guidelines: Awareness vs. Adherence. *SSRN Electronic Journal*, *July*.

Agha, L., Ericson, K. M., Geissler, K. H., & Rebitzer, J. B. (2022). Team Relationships and Performance: Evidence from Healthcare Referral Networks. *Management Science*, *68*(5), 3735–3754.

Agha, L., & Molitor, D. (2018). The Local Influence of Pioneer Investigators on Technology Adoption: Evidence from New Cancer Drugs. *The Review of Economics and Statistics*, *100*(1), 29–44.

Agha, L., & Zeltzer, D. (2022). Drug Diffusion through Peer Networks: The Influence of Industry Payments. *American Economic Journal: Economic Policy*, *14*(2), NBER working paper 26338.

Ahammer, A., & Schober, T. (2020). Exploring variations in health-care expenditures—What is the role of practice styles? *Health Economics (United Kingdom)*, *29*(6), 683–699.

Aksin, Z., Deo, S., Jonasson, J. O., & Ramdas, K. (2021). Learning from many: Partner exposure and team familiarity in fluid teams. *Management Science*, *67*(2), 854–874.

Alexander, D. (2015). Does Physician Pay Affect Procedure Choice and Patient Health? Evidence from Medicaid C-section Use. *Federal Reserve Bank of Chicago*, *WP 2017-07*.

Altman, M. R., Oseguera, T., Mclemore, M. R., Kantrowitz-gordon, I., Franck, L. S., & Lyndon, A. (2019). Social Science & Medicine Information and power: Women of color's experiences interacting with healthcare providers in pregnancy and birth. *Social Science & Medicine*, *238*(June), 112491.

Amaral-Garcia, S., Nardotto, M., Propper, C., & Valletti, T. (2022). Mums Go Online: Is the Internet Changing the Demand for Healthcare? *Review of Economics and Statistics*, *104*(6), 1157–1173.

Andrews, I., & Stock, J. H. (2018). Weak Instruments and What To Do About Them. *Methods Lectures*, 1–6.

Attanasio, L. B., Kozhimannil, K. B., & Kjerulff, K. H. (2018). Factors influencing women's perceptions of shared decision making during labor and delivery: Results from a large-scale cohort study of first childbirth. In *Patient Education and Counseling* (Vol. 101, Issue 6, pp. 1130–1136).

Avdic, D., Ivets, M., Lagerqvist, B., & Sriubaite, I. (2023). Providers, peers and patients. How do physicians' practice environments affect patient outcomes? *Journal of Health Economics*, *89*(February), WP 2021/01.

Avgerinos, E., & Gokpinar, B. (2017). Team familiarity and productivity in cardiac surgery operations: The effect of dispersion, bottlenecks, and task complexity. *Manufacturing \& Service Operations Management*, *19*(1), 19–35.

Barili, E., Bertoli, P., & Grembi, V. (2021). Fee equalization and appropriate healthcare. *Economics and Human Biology*, *41*.

Barrenho, E., Miraldo, M., Gautier, E., Propper, C., & Rose, C. (2023). Innovation Diffusion among Coworkers : Evidence from Senior Doctors. *CEPR Discussion Paper 15515*.

Barros, J. B., Goulart, A. C., Alencar, A. P., Lotufo, P. A., & Benseno, I. M. (2013). The influence of the day of the week of hospital admission on the prognosis of stroke patients. *Cadernos de Saúde Pública*, *29*(4), 769–777.

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. *American Economic Journal: Applied Economics*, *6*(2), 231–259.

Becker, D. J. (2007). Do Hospitals Provide Lower Quality Care on Weekends? *Health Services Research*, *42*(4), 1589–1612.

Berez, J., David, G., Howard, D. H., & Neuman, M. D. (2018). Does bad news travel faster? On the determinants of medical technology abandonment. *Journal of Human Capital*, *12*(4), 569–603.

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 Economics*, *29*(S1), 30–46.

Betran, A. P., Ye, J., Moller, A. B., Souza, J. P., & Zhang, J. (2021). Trends and projections of caesarean section rates: Global and regional estimates. *BMJ Global Health*, *6*(6), 1–8.

Betrán, A. P., Ye, J., Moller, A. B., Zhang, J., Gülmezoglu, A. M., & Torloni, M. R. (2016). The increasing trend in caesarean section rates: Global, regional and national estimates: 1990-2014. *PLoS ONE*, *11*(2), 1–12.

Borra, C., González, L., & Sevilla, A. (2016). Birth timing and neonatal health. *American Economic Review*, *106*(5), 329–332.

Borra, C., González, L., & Sevilla, A. (2019). The impact of scheduling birth early on infant health. *Journal of the European Economic Association*, *17*(1), 30–78.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*(1), 41–55.

Branco, D., Carrillo, B., Fang, D., & Iglesias, W. (2023). *Expertise Overlap and Team Productivity: Evidence from the Hospital Industry*.

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

Burke, M. A., Fournier, G. M., & Prasad, K. (2003). *Physician Social Networks and Geographical Variation in Medical Care *. 850*. https://www.brookings.edu/wp-content/uploads/2016/06/07healthcare_burke.pdf

Card, D., Fenizia, A., & Silver, D. (2023). the Health Impacts of Hospital Delivery Practices. *American Economic Journal: Economic Policy*, *15*(2), 42–81.

Cardador, M. T., Hill, P. L., & Salles, A. (2022). Unpacking the Status-Leveling Burden for Women in Male-Dominated Occupations. *Administrative Science Quarterly*, *67*(1), 237–284.

Chan, D. C. (2016). Teamwork and moral hazard: Evidence from the emergency department. *Journal of Political Economy*, *124*(3), 734–770.

Chan, D. C. (2021). Influence and Information in Team Decisions: Evidence from Medical Residency. *American Economic Journal: Economic Policy*, *13*(1), 106–137.

Chan, D. C., Gentzkow, M., & Yu, C. (2022). Selection with Variation in Diagnostic Skill: Evidence from Radiologists. *Quarterly Journal of Economics*, *137*(2), 729–783.

Chandra, A., & Staiger, D. O. (2007). Productivity spillovers in healthcare: Evidence from the treatment of heart attacks. *Journal of Political Economy*, *115*(1), 103–140.

Chandra, A., & Staiger, D. O. (2020). Identifying sources of inefficiency in healthcare. *Quarterly Journal of Economics*, *135*(2), 785–843.

Chen, C.-S., Liu, T.-C., Chen, B., & Lin, C.-L. (2014). The failure of financial incentive? The seemingly inexorable rise of cesarean section. *Social Science & Medicine*, *101*, 47–51.

Chen, Y. (2021). Team- Specific Human Capital and Team Performance: Evidence from Doctors. *American Economic Review*, *111*(12), 3923–3962.

Clarke, D. (2021). *RWOLF2: Stata module to calculate Romano-Wolf stepdown p-values for multiple hypothesis testing*. https://ideas.repec.org/c/boc/bocode/s458970.html

Cookson, G., & Laliotis, I. (2018). Promoting normal birth and reducing caesarean section rates: An evaluation of the Rapid Improvement Programme. *Health Economics*, *27*(4), 675–689.

Costa-Ramón, A., Kortelainen, M., Rodríguez-González, A., & Sääksvuori, L. (2021). The Long-Run Effects of Cesarean Sections. *Journal of Human Resources*.

Costa-Ramón, A., Rodríguez-González, A., Serra-Burriel, M., & Campillo-Artero, C. (2018). It's about time: Cesarean sections and neonatal health. *Journal of Health Economics*, *59*, 46–59.

Costa, F., Nunes, L., & Sanches, F. M. F. (2022). How to Attract Physicians to Underserved Areas? Policy Recommendations from a Structural Model. *The Review of Economics and Statistics*, 1–45.

Cuddy, E., & Currie, J. (2023). Rules vs. Discretion: Treatment of mental illness in U.S. adolescents. *Journal of Political Economy (Forthcoming)*.

Currie, J., & Macleod, W. B. (2017). Diagnosing expertise: Human capital, decision making, and performance among physicians. *Journal of Labor Economics*, *35*(1), 1–43.

Currie, J., & MacLeod, W. B. (2008). First do no harm? Tort reform and birth outcomes. *Quarterly Journal of Economics*, *123*(2), 795–830.

Currie, J., MacLeod, W. B., & Van Parys, J. (2016). Provider practice style and patient health outcomes: The case of heart attacks. *Journal of Health Economics*, *47*, 64–80.

Cutler, D., Skinner, J. S., Stern, A. D., & Wennberg, D. (2019). Physician Beliefs and Patient Preferences: A New Look at Supplier-Induced Demand. *American Economic Journal: Economic Policy*, *11*(1), 192–221.

Davidson, R., Roberts, S. E., Wotton, C. J., & Goldacre, M. J. (2010). Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: Evidence from the Oxford record linkage study. *BMC Pulmonary Medicine*, *10*, 2–9.

De Elejalde, R., & Giolito, E. (2021). A demand-smoothing incentive for cesarean deliveries. *Journal of Health Economics*, *75*.

De Giorgi, G., Frederiksen, A., & Pistaferri, L. (2019). Consumption Network Effects. *American Historical Review*, *124*(2), 130–163.

De Giorgi, G., Pellizzari, M., & Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, *2*(2), 241–275.

De Muylder, X. (1993). Caesarian sections in developing countries: Some considerations. *Health Policy and Planning*, *8*(2), 101–112.

De Oliveira, V. H., Lee, I., & Quintana-Domeque, C. (2022). The effect of increasing Women's autonomy on primary and repeated caesarean sections in Brazil. *Health Economics*, *31*(8), 1800–1804.

Dickert-Conlin, S., & Chandra, A. (1999). Taxes and the timing of births. *Journal of Political Economy*, *107*(1), 161–177.

Domingues, R. M. S. M., Dias, M. A. B., Nakamura-Pereira, M., Torres, J. A., D'Orsi, E., Pereira, A. P. E., Schilithz, A. O. C., & Leal, M. do C. (2014). Process of decision-making regarding the mode of birth in Brazil: from the initial preference of women to the final mode of birth. *Cadernos de Saude Publica*, *30*(SUPPL1), 1–16.

Doyle, J. J., Ewer, S. M., & Wagner, T. H. (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of Health Economics*, *29*(6), 866–882.

Doyle, J. J., & Staiger, B. (2022). Physician Group Influences on Treatment Intensity and Health: Evidence from Physician Switchers. *NBER Working Paper*, *29613*. https://www.nber.org/papers/w29613

Epstein, A. J., & Nicholson, S. (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*.

Epstein, A. J., Nicholson, S., & Asch, D. A. (2016). The production of and market for new physicians' skill. *American Journal of Health Economics*, *2*(1), 41–65.

Evangelista, P. A., Barreto, S. M., & Guerra, H. L. (2008). Original Article Hospital Admission and Hospital Death Associated to Ischemic Heart Diseases at the National Health System (SUS). *Arquivos Brasileiros de Cardiologia*, *90*(2), 119–126.

Fabbri, D., Monfardini, C., Castaldini, I., & Protonotari, A. (2016). Cesarean section and the manipulation of exact delivery time. *Health Policy*, *120*(7), 780–789.

Facchini, G. (2022). Forgetting-by-not-doing: The case of surgeons and cesarean sections. *Health Economics (United Kingdom)*, *31*(3), 481–495.

Fadlon, I., & Van Parys, J. (2020). Primary care physician practice styles and patient care: Evidence from physician exits in medicare. *Journal of Health Economics*, *71*, 102304.

Ferraro, J., Khare, S., & Acosta, A. (2021). *Physician convenience and cesarean section delivery*.

Foo, P. K., Lee, R. S., & Fong, K. (2017). Physician prices, hospital prices, and treatment choice in labor and delivery. *American Journal of Health Economics*, *3*(3), 422–453.

Gans, J., Leigh, A., & Varganova, E. (2007). Minding the shop: The case of obstetrics conferences. *Social Science & Medicine*, *65*(7), 1458–1465.

Gans, J. S., & Leigh, A. (2009). Born on the first of July: An (un)natural experiment in birth timing. *Journal of Public Economics*, *93*(1–2), 246–263.

Gijsen, R., Hukkelhoven, C. W. P. M., Schipper, C. M. A., Ogbu, U. C., de Bruin-Kooistra, M., & Westert, G. P. (2012). Effects of hospital delivery during off-hours on perinatal outcome in several subgroups: a retrospective cohort study. *BMC Pregnancy and Childbirth*, *12*, 27–33.

Ginja, R., Riise, J., Willage, B., & Willén, A. (2022). Does Your Doctor Matter? Doctor Quality and Patient Outcomes. *SSRN Electronic Journal*, *June*.

Gong, Q. (2018). Physician Learning and Treatment Choices: Evidence from Brain Aneurysms. *Working Paper*. https://sites.google.com/site/qgongecon/

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.

Gould, J. B., Qin, C., Marks, A. R., & Chavez, G. (2003). Neonatal Mortality in Weekend vs Weekday Births. *Journal of the American Medical Association*, *289*(22), 2958–62.

Gowrisankaran, G., Joiner, K., & Léger, P. T. (2022). Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments. *Management Science*, 1–61.

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

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological Assessment*, *12*(1), 19.

Gruber, J., Kim, J., & Mayzlin, D. (1999). Physician fees and procedure intensity: The case of cesarean delivery. *Journal of Health Economics*, *18*(4), 473–490.

Gruber, J., & Owings, M. (1996). Physician Financial Incentives and Cesarean Section

Delivery. *The RAND Journal of Economics*, *27*(1), 99.

Grytten, J., Monkerud, L., & Sørensen, R. (2012). Adoption of Diagnostic Technology and Variation in Caesarean Section Rates: A Test of the Practice Style Hypothesis in Norway. *Health Services Research*, *47*(6), 2169–2189.

Grytten, J., & Sørensen, R. (2003). Practice variation and physician-specific effects. *Journal of Health Economics*, *22*(3), 403–418.

Håkansson, S., & Källén, K. (2003). Caesarean section increases the risk of hospital care in childhood for asthma and gastroenteritis. *Clinical and Experimental Allergy*, *33*(6), 757–764.

Hamilton, K. E. S., Redshaw, M. E., & Tarnow-Mordi, W. (2007). Nurse staffing in relation to risk-adjusted mortality in neonatal care. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *92*(2), F99–F103.

Hansen, A. K., Wisborg, K., Uldbjerg, N., & Henriksen, T. B. (2008). Risk of respiratory morbidity in term infants delivered by elective caesarean section: Cohort study. *Bmj*, *336*(7635), 85–87.

Hockenberry, J. M., & Helmchen, L. A. (2014). The nature of surgeon human capital depreciation. *Journal of Health Economics*, *37*(1), 70–80.

Hong, J., Kang, H., Yi, S.-W., Han, Y., Nam, C., Gombojav, B., & Ohrr, H. (2006). A comparison of perinatal mortality in Korea on holidays and working days. *BJOG*, *113*(11), 1235--1238.

Huang, S., & Ullrich, H. (2023). Physician Effects in Antibiotic Prescribing: Evidence from Physician Exits. *SSRN Electronic Journal*, 1958.

Jachetta, C. (2016). *Cesarean Sections and Later Child Health Outcomes* (Issue 2000). University of Illinois.

Jacobson, M., Kogelnik, M., & Royer, H. (2021). Holiday, just one day out of life: Birth timing and postnatal outcomes. *Journal of Labor Economics*, *39*(S2), S651–S702.

Jensen, E. A., & Lorch, S. A. (2017). Association between Off-Peak Hour Birth and Neonatal Morbidity and Mortality among Very Low Birth Weight Infants. *Journal of Pediatrics*, *186*, 41-48.e4.

Johnson, E. J. (1988). Expertise and decision under uncertainty: Performance and process. *The Nature of Expertise*, 209–228.

Johnson, E. M., & Rehavi, M. M. (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, *8*(1), 115–141.

Keeler, E., & Fok, T. (1996). Equalizing physician fees had little effect on cesarean rates. *Medical Care Research and Review*, *53*(4), 465–471.

Kozhimannil, K. B., Graves, A. J., Ecklund, A. M., Shah, N., Aggarwal, R., & Snowden, J. M. (2018). Cesarean delivery rates and costs of childbirth in a state Medicaid program after implementation of a blended payment policy. *Medical Care*, *56*(8), 658–664.

Kristensen, K., & Henriksen, L. (2016). Cesarean section and disease associated with immune function. *Journal of Allergy and Clinical Immunology*, *137*(2), 587–590.

Kwok, J. H. (2019). How Do Primary Care Physicians Influence Healthcare? Evidence on Practice Styles and Switching Costs from Medicare. In *SSRN Electronic Journal* (Job Market Paper, Issue July).

Lancet, T. (2000). Caesarean section on the rise. In *The Lancet* (Vol. 356, Issue 9243, p. 1697). Elsevier.

Leal, M. do C., Gama, S. G. N. da, Pereira, A. P. E., Pacheco, V. E., Carmo, C. N. do, & Santos, R. V. (2017). The color of pain: racial iniquities in prenatal care and childbirth in Brazil. *Cadernos de Saude Publica*, *33*, e00078816.

Lefevre, M. (2014). Physician Induced Demand for C-Sections: Does the Convenience Incentive Matter? In *Health, Econometrics and Data Group* (14/08).

Leivas, P. H. S. (2017). *Três ensaios em economia hospitalar. Weekend effect nos atendimentos de urgência no Brasil: O caso do Infarto Agudo do Miocárdio*. Pontifícia Universidade Católica do Rio Grande do Sul.

Leuven, E., & Rønning, M. (2016). Classroom Grade Composition and Pupil Achievement. *Economic Journal*, *126*(593), 1164–1192.

Levin, J. (2006). *A Influência das Políticas de Saúde nos Indicadores Gerados pelo Sistema de Informações Hospitalares do SUS*. Universidade do Estado do Rio de Janeiro.

Lo, J. C. (2003). Patients' attitudes vs. physicians' determination: Implications for cesarean sections. *Social Science & Medicine*, *57*(1), 91–96.

Lo, J. C. (2008). Financial incentives do not always work—an example of cesarean sections in Taiwan. *Health Policy*, *88*(1), 121–129.

Lundborg, P., James, S., Lagerqvist, B., & Vikström, J. (2021). Learning-by-Doing and Productivity Growth Among High-Skilled Workers: Evidence from the Treatment of Heart Attacks. *SSRN Electronic Journal*, *14744*.

Lyndon, A., Lee, H. C., Gay, C., Gilbert, W. M., Gould, J. B., & Lee, K. A. (2015). Effect of time of birth on maternal morbidity during childbirth hospitalization in California. *American Journal of Obstetrics and Gynecology*, *213*(5), 705-e1.

Machado, D. da S., Lelis, D. A. S. de, & Clark, G. (2022). Tabela de Procedimentos do SUS à luz da ordem econômica: ausência de correção inflacionária da remuneração das Santas Casas no âmbito da saúde pública. *Revista Estudos Institucionais*, *8*(3), 481–506.

Magid, D. J., Wang, Y., Herrin, J., McNamara, R. L., Bradley, E. H., Curtis, J. P., Pollack, C. V., French, W. J., Blaney, M. E., & Krumholz, H. M. (2005). Relationship between time of day, day of week, timeliness of reperfusion, and in-hospital mortality for patients with acute ST-segment elevation myocardial infarction. *Journal of the American Medical Association*, *294*(7), 803–812.

Maibom, J., Sievertsen, H. H., Simonsen, M., & Wüst, M. (2021). Maternity ward crowding, procedure use, and child health. *Journal of Health Economics*, *75*.

Manski, C. F. (1993). Identification of endogenous social effects the reflection problem. *Review of Economic Studies*, *60*(3), 531–542.

Manski, C. F. (2017). Improving Clinical Guidelines and Decisions under Uncertainty Charles. *NBER Working Paper Series*, 61. http://www.nber.org/papers/w23915

Manski, C. F. (2018). Reasonable patient care under uncertainty. In *Health Economics (United Kingdom)* (Vol. 27, Issue 10). Princeton University Press.

Manski, C. F. (2019). *Patient care under uncertainty*. Princeton University Press.

Marquardt, K. (2022). Physician Practice Style for Mental Health Conditions: The Case of ADHD. *SSRN Electronic Journal*, *August 2021*.

Marquardt, K. (2023). *Mis(sed) Diagnosis : Physician Decision Making and ADHD*. 1–63.

McLemore, M. R., Altman, M. R., Cooper, N., Williams, S., Rand, L., & Franck, L. (2018). Healthcare experiences of pregnant, birthing and postnatal women of color at risk for preterm birth. *Social Science and Medicine*, *201*(September 2017), 127–135.

Mello, F. C. M., Silva, J. L. da, Oliveira, W. A. de, Prado, R. R. do, Malta, D. C., & Silva, M. A. I. (2017). The practice of bullying among Brazilian schoolchildren and associated factors , National School Health Survey 2015. *Ciência & Saúde Coletiva*, *22*(9), 2939–2948.

Melo, C., & Filho, N. M. (2021). *Birth timing manipulation during Carnival in Brazil*.

Melo, C., & Menezes-Filho, N. (2023). The effects of a national policy to reduce c-sections in Brazil. *Health Economics*, *32*(2), 501–517.

Melo, L. (2021). Restricting the timing of elective cs: Evidence from Brazil. *Estudos Economicos*, *51*(2), 245–283.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy*, *10*(1), 326–356.

Montiel Olea, J. L., & Pflueger, C. (2013). A Robust Test for Weak Instruments. *Journal of Business and Economic Statistics*, *31*(3), 358–369.

Moore, H. C., De Klerk, N., Holt, P., Richmond, P. C., & Lehmann, D. (2012). Hospitalisation for bronchiolitis in infants is more common after elective caesarean delivery. *Archives of Disease in Childhood*, *97*(5), 410–414.

Mullainathan, S., & Obermeyer, Z. (2022). Diagnosing physician error: A machine learning approach to low-value healthcare. *The Quarterly Journal of Economics*, *137*(2), 679–727.

Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *Review of Economic Studies*, *74*(3), 897–935.

Nicoletti, C., Salvanes, K. G., & Tominey, E. (2018). The family peer effect on mothers' labor supply. *American Economic Journal: Applied Economics*, *10*(3), 206–234.

Nobrega, J. K. (2015). *What is Pushing Brazil Not to Push?* University of Illinois at Chicago.

Okeke, E. N., & Chari, A. V. (2018). Healthcare at birth and infant mortality: Evidence from nighttime deliveries in Nigeria. *Social Science and Medicine*, *196*(November 2017), 86–95.

Palmer, W. L., Bottle, A., & Aylin, P. (2015). Association between day of delivery and obstetric outcomes. *BMJ*, *351*.

Patnam, M. (2015). *Corporate networks and peer effects in firm policies*. https://www.econ.cam.ac.uk/conf/networks-docs/corporate_mpatnam.pdf

Pflueger, C. E., & Wang, S. (2015). A robust test for weak instruments in Stata. *Stata Journal*, *15*(1), 216–225.

Phelps, C. E. (1993). *Variations in medical practice use: causes and consequences*.

Phelps, C. E. (2000). Chapter 5 Information diffusion and best practice adoption. *Handbook of Health Economics*, *1*(PART A), 223–264.

Phelps, C. E., Mooney, C., Mushlin, A. I., Handy, B., & Perkins, N. (1994). Doctors have styles, and they matter! *University of Rochester Working Paper*.

Pilvar, H., & Yousefi, K. (2021). Changing physicians' incentives to control the C-section rate: Evidence from a major healthcare reform in Iran. *Journal of Health Economics*, *79*.

Robinson, S., Royer, H., & Silver, D. (2023). Geographic Variation in Cesarean Sections in the United States: Trends, Correlates, and Other Interesting Facts. *NBER Working Paper*. http://www.nber.org/papers/w31871

Roduit, C., Scholtens, S., De Jongste, J. C., Wijga, A. H., Gerritsen, J., Postma, D. S., Brunekreef, B., Hoekstra, M. O., Aalberse, R., & Smit, H. A. (2009). Asthma at 8 years of age in children born by caesarean section. *Thorax*, *64*(2), 107–113.

Salam, M. T., Margolis, H. G., McConnell, R., McGregor, J. A., Avol, E. L., & Gilliland, F. D. (2006). Mode of Delivery Is Associated With Asthma and Allergy Occurrences in Children. *Annals of Epidemiology*, *16*(5), 341–346.

Sarsons, H. (2017). Interpreting Signals in the Labor Market: Evidence from Medical Referrals. In *Job Market Paper*.

Scheffer, M., Cassenote, A., Guilloux, A. G. A., Biancarelli, A., Miotto, B. A., & Mainardi, G. M. (2018). *Demografia Médica No Brasil 2018*.

Scheffer, M., Guilloux, A. G. A., Miotto, B. A., & Almeida, C. de J. (2023). *Demografia Médica no Brasil 2023*. https://amb.org.br/

Schulkind, L., & Shapiro, T. M. (2014). What a difference a day makes: Quantifying the effects of birth timing manipulation on infant health. *Journal of Health Economics*, *33*(1), 139–158.

Silver, D. (2021). Haste or Waste? Peer Pressure and Productivity in the Emergency Department. *The Review of Economic Studies*, *88*(3), 1385–1417.

Slaughter-Acey, J. C., Sneed, D., Parker, L., Keith, V. M., Lee, N. L., & Misra, D. P. (2019). Skin Tone Matters: Racial Microaggressions and Delayed Prenatal Care. *American Journal of Preventive Medicine*, *57*(3), 321–329.

Song, Y., Skinner, J., Bynum, J., Sutherland, J., Wennberg, J. E., & Fisher, E. S. (2010). Regional Variations in Diagnostic Practices. *New England Journal of Medicine*, *363*(1), 45–

53.

Sosnaud, B. (2021). Cross-State Differences in the Processes Generating Black–White Disparities in Neonatal Mortality. *Demography*.

Stecher, C. (2023). Productivity Gains from Shared Experience. *American Journal of Health Economics*, *9*(2).

Stoye, G. (2022). *The distribution of doctor quality: evidence from cardiologists in England* (22/30). 22/30.

Thavagnanam, S., Fleming, J., Bromley, A., Shields, M. D., & Cardwell, C. R. (2008). A meta-analysis of the association between Caesarean section and childhood asthma. *Clinical and Experimental Allergy*, *38*(4), 629–633.

Tollånes, M. C., Moster, D., Daltveit, A. K., & Irgens, L. M. (2008). Cesarean Section and Risk of Severe Childhood Asthma: A Population-Based Cohort Study. *Journal of Pediatrics*, *153*(1), 112–117.

Treacy, L., Bolkan, H. A., & Sagbakken, M. (2018). Distance, accessibility and costs. Decision-making during childbirth in rural Sierra Leone: A qualitative study. *PLoS ONE*, *13*(2), 1–17.

Trogdon, J. G., Nonnemaker, J., & Pais, J. (2008). Peer effects in adolescent overweight. *Journal of Health Economics*, *27*(5), 1388–1399.

Tsugawa, Y., Jha, A. K., Newhouse, J. P., Zaslavsky, A. M., & Jena, A. B. (2017). Variation in physician spending and association with patient outcomes. *JAMA Internal Medicine*, *177*(5), 675–682.

Tu, P. (2017). *Quasi-experimental Evidence of Physician Effects* (Job Mark. Pap.; Issue November). http://scholar.harvard.edu/petertu

Valdes, E. G. (2021). Examining Cesarean Delivery Rates by Race: a Population-Based Analysis Using the Robson Ten-Group Classification System. *Journal of Racial and Ethnic Health Disparities*, *8*(4), 844–851.

Van Parys, J. (2016). Variation in physician practice styles within and across emergency departments. *PLoS ONE*, *11*(8), 1–19.

Vedam, S., Stoll, K., Taiwo, T. K., Rubashkin, N., Cheyney, M., Strauss, N., McLemore, M., Cadena, M., Nethery, E., Rushton, E., Schummers, L., & Declercq, E. (2019). The Giving Voice to Mothers study: Inequity and mistreatment during pregnancy and childbirth in the United States. *Reproductive Health*, *16*(1), 1–18.

Wallace, J. E. (2014). Gender and supportive co-worker relations in the medical profession. *Gender, Work and Organization*, *21*(1), 1–17.

Yang, M., Lien, H. M., & Chou, S. Y. (2014). Is there a physician peer effect? Evidence from new drug prescriptions. *Economic Inquiry*, *52*(1), 116–137.

Zampieri, F. G., Lisboa, T. C., Correa, T. D., Bozza, F. A., Ferez, M., Fernandes, H. S., Japiassú, A. M., Verdeal, J. C. R., Carvalho, A. C. P., Knibel, M. F., Mazza, B. F., Colombari, F., Vieira, J. M., Viana, W. N., Costa, R., Godoy, M. M., Maia, M. O., Caser, E. B., Salluh, J. I. F., & Soares, M. (2018). Role of organisational factors on the 'weekend effect'in critically ill patients in Brazil: a retrospective cohort analysis. *BMJ Open*, *8*(1), e018541.

Zapf, M. A. C., Kothari, A. N., Markossian, T., Gupta, G. N., Blackwell, R. H., Wai, P. Y., Weber, C. E., Driver, J., & Kuo, P. C. (2015). The weekend effect in urgent general operative procedures. *Surgery*, *158*(2), 508–514.