The Unexpected Costs of Expertise: Evidence from Highly Specialized Physicians [*]

Yi Cheng [†]

Job Market Paper

See the latest version here

Abstract

High U.S. spending on health care is commonly attributed to its intensity of specialized, high-tech medical care. A growing body of research focuses on physicians whose medical decisions shape treatment intensity, costs, and patient outcomes. Often overlooked in this research is the assignment of physician skills to patient conditions, which may strongly affect health outcomes and productivity. This matching may be especially important in the case of hospital admissions as high-frequency fluctuations in patient flow make it challenging to maintain effective matches between the best-suited physicians and their patients. This paper focuses on hospitals' responses to demand shocks induced by unscheduled high-risk admissions. I show that these demand shocks result in physician–patient mismatches when hospitals are congested. Specifically, highly specialized physicians who are brought in to treat unscheduled high-risk admissions also treat previously admitted lower-risk patients. This leads to increased treatment intensity for lower-risk patients, which I attribute to persistence in physician practice style. Despite the greater treatment intensity, I find no detectable improvement in health outcomes, which *prima facie* could be viewed as waste. However, the mismatches observed only at high congestion levels more likely reflect hospitals' careful assessment of costs and benefits when assigning physicians to patients – maintaining preferred physician–patient matching can be particularly costly when congestion is high. My findings highlight the need to consider both heterogeneity within patient and physician type, and furthermore show how the common phenomenon of demand uncertainty can promote mismatch between these types.

[*]I thank Douglas Almond, Micheal Best, Sandra Black, Janet Currie, Joseph Doyle, Francois Gerard, Wojciech Kopczuk, Bentley MacLeod, Suresh Naidu, Adam Sacarny, Jeffrey Shrader, Ashley Swanson, Miguel Urquiola, my fellow Ph.D. classmates, and participants of the Columbia Applied Micro Research Methods Colloquium and Applied Micro & Labor Workshop for helpful comments. I also thank doctors Brett Anderson and Christopher Almond for helpful discussions regarding neonatal care for high-risk infants, and Solveig Edlund for proofreading. All errors are my own.

[†]Department of Economics, Columbia University. Email: yi.cheng@columbia.edu

1 Introduction

At 17.1% of its GDP, the United States spends twice the OECD average on health care. Given its unsatisfactory average health outcomes, there is substantial interest in the effectiveness of U.S. spending on health care and whether providers can reduce spending without compromising quality and health outcomes. Existing research has yet to reach a consensus on the magnitude and sources of waste (Fisher et al., 2003a,b; Skinner and Fisher, 2010; Almond and Doyle, 2011; Chandra et al., 2011; Doyle et al., 2017). In part, this is due to empirical challenges in measuring productivity and identifying unproductive spending in the health care system.

Although payments to physicians only constitute a small fraction of the aggregate health care spending, physicians' medical decisions clearly shape care utilization and patient outcomes (Phelps, 2016). A growing body of research shows that increasing specialization leads to large variation in skills and practice styles among medical professionals, and that specialized physicians tend to adopt more intensive practice styles.[1] However, empirical findings on whether physician specialists, or physicians with more intensive styles, provide higher quality care are mixed (Doyle et al., 2010; Baicker and Chandra, 2004; Currie et al., 2016; Currie and MacLeod, 2017; Molitor, 2018; Fadlon and Van Parys, 2019). Existing studies usually evaluate productivity focusing on a constant, physician-specific measure. It is often overlooked that productivity may vary within physicians, depending on the type of patients they are treating. Therefore, the matching between physicians' skills and patients' conditions ("skill–task matching") can affect both care utilization and patient outcomes.

This paper examines skill–task matching in hospital admissions. Assessing the impact of physician–patient assignments is challenging. On the one hand, high levels of specialization and large variations in practice styles across medical professionals would increase the gains from matching the best-suited physicians to patients. On the other hand, due to the variability and unpredictability in patient flow, maintaining good physician–patient matchings for every admission can be costly, and even outweigh the benefits of physicians' specialized skills. Previous studies have recognized that fluctuations in patient flow incur costs to both hospitals and patients, and policies have attempted to relieve congestion in order to improve care quality (Hughes and McGuire, 2003; Evans and Kim, 2006; Hoot and

[1] In this study, I do not differentiate between physicians' skills and practice styles due to the fact that they are highly correlated. Therefore, I use the terms "practice style", "skill", and "expertise" interchangeably.

Aronsky, 2008; Sharma et al., 2008; Shen and Hsia, 2015; Hoe, 2018; Marks and Choi, 2018). However, relatively unexplored is whether the need of matching physician type to patients is an important source of costs. To my knowledge, this paper is the first to provide empirical evidence on skill–task matching stemming from demand uncertainty in the health care sector.

In particular, I analyze how short-term demand fluctuations induced by unscheduled high-risk admissions affect health care production in hospitals. I pay special attention to whether hospitals develop differential responses depending on the level of costs or difficulties in achieving good physician–patient matchings, and how in turn these responses affect health care production. The level of hospital congestion serves as a proxy for the costs of matching in this study. Finding a physician who specializes in treating a certain condition is relatively easy when many physicians are available. But when hospitals become congested, achieving good matchings for every patient may become difficult: more physicians are occupied and the skill range of available physicians becomes more limited.

Using New York City hospital discharge micro data, I focus on newborns. Childbirth is the most common reason for hospitalization in the United States (HCUP, 2015) and at-risk newborns disproportionately drive the high aggregate spending on neonatal care (Torio and Moore, 2016). Hospital discharge records provide rich information on physicians and patients, treatment decisions, and health outcomes. Additionally, patients' arrivals and assignments to physicians are explicitly recorded in the high-frequency micro data, allowing for a detailed study of skill–task matching and its effects on health care production. Furthermore, effects on newborn health can lead to long term impacts later in life, such as educational attainment, adult disability, and labor market outcomes (Currie, 2009; Bharadwaj et al., 2013; Figlio et al., 2014; Elder et al., 2019).

Birth weight is the most commonly used metric of newborn health both in the literature and in medical practice. Newborns weighing less than 1500 grams ("very low birth weight") require immediate and intensive neonatal care. The precise timing of vaginal deliveries is hard for hospitals to predict. Hence, vaginally-delivered very low birth weight births may serve as demand shocks to hospitals. In this study, I refer to vaginally-delivered very low birth weight infants as "high-risk" unscheduled admissions. Using an event study framework, I find that hospitals summon physicians with more intensive practice styles who specialize in treating high-risk newborns upon unscheduled high-risk admissions. Critical for my purpose, these highly specialized physicians who are called in also treat previously admitted newborns ("incumbent newborns"). This spillover effect is especially pronounced when hospitals are congested, creating exogenous variation in the typical physician–patient matching.

I demonstrate that newborns admitted *prior to* unscheduled high-risk admissions and newborns not affected by any demand shocks do not differ in observables at admission, which supports the exogeneity of my demand shocks. When hospitals are congested, lower-risk newborns admitted just before unscheduled high-risk admissions are more likely to be treated by highly specialized physicians, leading to increased treatment intensity. Despite being treated more intensively, little improvement in patient outcomes is seen. This suggests low, even zero, marginal returns to care utilization. Many studies have established that specialists and their intensive practice styles can benefit high-risk patients (Currie et al., 2016; Currie and MacLeod, 2017; Doyle, 2018). Results in this study, however, suggest that physician productivity is patient-dependent: physicians who specialize in treating high-risk patients may provide low-return care when treating lower-risk patients. Notably, this low return is found for newborns weighing between 1500 - 2500 grams who are "mid-risk", i.e. excluding normal birth weight infants. These findings highlight the importance of matching physicians' skills to patients' conditions in health care production and point to the potential costs associated with physician experts beyond physician payments.

Prima facie, the low productivity resulting from physician–patient mismatch may appear purely wasteful. However, it is worth emphasizing that such low productivity is *only* detectable when hospitals are congested and matching the best-suited physicians to patients is costly. This finding is consistent with predictions from a stylized model: optimal decisions depend on the relative magnitudes of costs and benefits associated with achieving good matchings; allowing a degree of mismatch can be optimal if the matching costs are high. Hence, the mismatch observed at high congestion levels may reflect hospitals' careful assessment of costs and benefits when assigning physicians to patients. Analyses of incumbent newborn characteristics also suggest that hospitals attempt to maintain good physician–patient matchings given the availability of physicians: among mid-risk incumbent infants, newborns with worse health conditions (although still healthier than the high-risk newborns) tend to be assigned to the highly specialized physicians. In more extreme cases, I find that the highly specialized physicians do not treat any incumbent newborns when the expected returns to their specialized skills are too low.

This paper contributes to the literature on physician productivity and health care production. The evidence on patient-dependent physician productivity presented in this paper provides a possible explanation for the lack of research consensus on how physicians' skills and treatment intensity affect patient outcomes: the productivity response is shaped by which subpopulation of patients are treated. In addition, by restricting empirical comparisons to be within hospitals, this study helps isolate the effect of physician practice on care utilization. This paper complements existing literature on regional or cross-hospital variation in medical spending by arguing that physicians' practice styles contribute to variations in spending among many other factors, such as differences in patient composition or facility quality (Baicker and Chandra, 2004; Chandra and Staiger, 2007; Doyle, 2011; Cutler et al., 2013; Doyle et al., 2015, 2017; Molitor, 2018).

This paper also contributes to non-health literature on production under specialization. While theories on specialization and coordination in labor and organizational economics exist, these studies mostly focus on the optimal level of specialization (Becker and Murphy, 1992; Garicano, 2000; Dessein and Santos, 2006; Fuchs and Garicano, 2010). Little is known about how firms respond to fluctuations in demand given the current level of specialization among their employees. This is a particularly common situation in day-to-day production decisions where labor inputs are costly to constantly adjust. This study helps to fill this gap in research by demonstrating empirically that hospitals evaluate carefully the associated benefits and costs when making job assignments to specialists under demand uncertainty.

The rest of the paper proceeds as follows. Section 2 provides background information and discusses the relevant literature. Section 3 describes the empirical strategy and the data. Section 4 presents results on hospitals' responses to unscheduled high-risk admissions. Section 5 reports the spillover effects of unscheduled high-risk admissions. Section 6 discusses potential mechanisms for the observed effects and their implications. Section 7 concludes.

2 Background

2.1 Physician Specialization and Practice Style

The expansion in human capital over time enables the increases in specialization. Along with other highly-skilled occupations, physicians are increasingly specialized. The number of physician specialists grew six times faster than the number of primary care doctors from 2005 to 2015 (Barbey et al., 2017). By the end of 2017, more than two-thirds of physicians had specialties outside primary care (AAMC, 2018). In the United States, highly specialized physicians are required to take intensive training post M.D. in order to handle complex patient conditions. For example, a physician needs to complete an additional three-year residency to be certified as a pediatrician and a second three-year neonatology fellowship to treat critically ill newborns as a neonatologist.

Physicians incorporate their own judgment and expertise when making medical decisions. The lack of comprehensive medical guidelines and the complexity of patient conditions lead to large variation in physicians' practice styles (Phelps, 2016). Empirical evidence has shown that specialists tend to utilize more intensive treatments (Doyle, 2018). Style differences can arise from skill differences, differences in the assessment of treatment efficacy and patient conditions that are shaped by past training and experiences, or other factors (Chandra et al., 2011). In the case of neonatal care, it's not hard to imagine that a neonatologist's practice style will differ from a general pediatrician's because of the additional fellowship training, just as economics PhD students are likely to approach economics questions differently from what they would do in college.

Previous studies usually measure physician practice styles using a physician-specific "fixed effect". It is challenging empirically to disentangle whether the high intensity practice observed among physician specialists is due to selection bias (i.e. highly specialized physicians usually treat more severe patients) or due to something inherent to physicians themselves. This paper utilizes a natural experiment to overcome selection bias and demonstrates that highly specialized physicians also provide more intensive care when treating lower-risk patients. Results in this paper provide evidence that physicians' practice styles indeed persist across patients and thereby influence the overall care utilization.

2.2 Productivity of Specialized Physicians

Specialists spend more. However, evidence on patient health impacts has been mixed. Baicker and Chandra (2004) find that states with more specialists have higher costs and lower care quality. Doyle (2018) provides evidence that heart failure patients receive more intensive treatments and are more likely to survive at one year when more cardiologists are available. Currie et al. (2016) show that physicians with better procedural skills provide more aggressive treatments, which benefits patients in the case of heart attacks. Currie and MacLeod (2017) assess physician productivity in both decision making and procedural skill. They show that improvement in each dimension can lead to better outcomes. Doyle et al. (2010) consider physician human capital and find that physicians from a higher ranking medical school have better decision making, which reduces costs without affecting patient outcomes.

This paper considers a less studied subject: what happens when highly specialized physicians treat lower-risk patients? I find that the more intensive treatments assigned by specialists do not generate detectable health improvements. While there has been evidence that high-risk patients are likely to benefit from specialists' intensive treatments, this paper adds to existing knowledge by showing that the productivity of highly specialized physicians are not universal but instead task-dependent. It further highlights the importance of considering skill–task matching in assessing physician productivity. Finally, it suggests a mechanism by which the mixed evidence on specialists' productivity may be reconciled.

2.3 Demand Fluctuation and Hospital Congestion

Hospitals face frequent demand fluctuations, which often result in congestion. Existing studies find that variability in demand is costly to hospitals (Baker et al., 2004). Studies evaluating the effects of hospital congestion *per se* on patients often find lower care quality and worse health outcomes (Evans and Kim, 2006; Bartel et al., 2011; Shen and Hsia, 2015; Hoe, 2018). Among studies that examine fluctuations in patient flow, Freedman (2016) explores congestion levels in neonatal intensive care units (NICU) and shows that physicians make NICU admission decisions based on bed availability: empty NICU beds increase NICU admission for marginally sick infants but have little or no effect for the sickest infants. Departing from Freedman (2016), I utilize fluctuations in NICU occupancy as a proxy for the level of difficulty in achieving good matchings between physicians and patients. In addition, this paper exploits fluctuations in hospital demand beyond generic patient volume. By taking into account variation in patient acuity that results from unscheduled high-risk admissions, this paper suggests that demand fluctuations can affect care utilization and patient outcomes through an under-recognized channel, i.e. by affecting the assignment of physicians to patients. This yields new implications for policies aiming at hospital congestion management.

3 Research Design and Data

3.1 Research Design

This study analyzes the production process in the professional industry particularly focusing on the matching of skills to tasks. In the case of hospital admission, patients are typically assigned to physicians based on perceived patient condition and physician expertise. It is commonly observed that high-risk patients are assigned to highly specialized physicians with intensive practice styles. Hence such selection bias usually impedes the empirical evaluation of productivity in the health care sector, or in any industries where better skills are overwhelmingly bundled with difficult tasks. This paper overcomes the selection bias by exploiting an exogenous variation in physician–patient matching and care utilization resulting from short-term demand fluctuations. When an unscheduled high-risk patient is admitted, hospitals frequently need to adjust resource allocation among previously admitted patients, including physician assignment, to accommodate the unexpected increases in care demand.

In this study, I investigate hospitals' responses to temporary demand shocks in the neonatal care sector. I first document the strategies hospitals adopt when an unexpected high-risk patient is admitted under an event study framework. To utilize the variation in resource allocation among previously admitted newborns induced by unexpected high-risk admissions, I compare newborns admitted just prior to unscheduled high-risk admissions (treated group) to those having little overlap with any unscheduled high-risk admissions (control group). If the unexpected demand shocks are quasi-random in time, newborns in the treated and control groups are expected to be comparable in all aspects upon birth admission. In this case, any differences in subsequent outcomes can be attributed to differences in changes of care provision induced by the unexpected high-risk newborn admissions. Although some may argue that patients tend to choose their physicians in the case of birth delivery, such selection happens mostly in the form of mothers choosing their obstetricians. Since the focus of this study is on newborn infants whose physicians are pediatricians, such selection is unlikely to occur. In the case of at-risk newborns who are admitted to NICU upon their births, it is even less likely that the parents will have any discretion to choose physicians.

I define unscheduled high-risk admissions to be vaginally-delivered very low birth weight newborns, noting first that birth weight has been shown as a good metric of newborn acuity and expected care utilization. In the medical definition, low birth weight is defined to be below 2,500 grams (5 pounds 8 ounces). Subcategories include very low birth weight, which is less than 1500 grams (3 pounds 5 ounces), and extremely low birth weight, which is less than 1000 grams (2 pounds 3 ounces). Medical diagnosis codes are assigned to each low birth weight categories and physicians use these birth weight cutoffs to make differential treatment decisions. Very low birth weight newborns utilize an extremely high amount of care resources, hence leading to large increases in demand at hospitals (Almond et al., 2010).

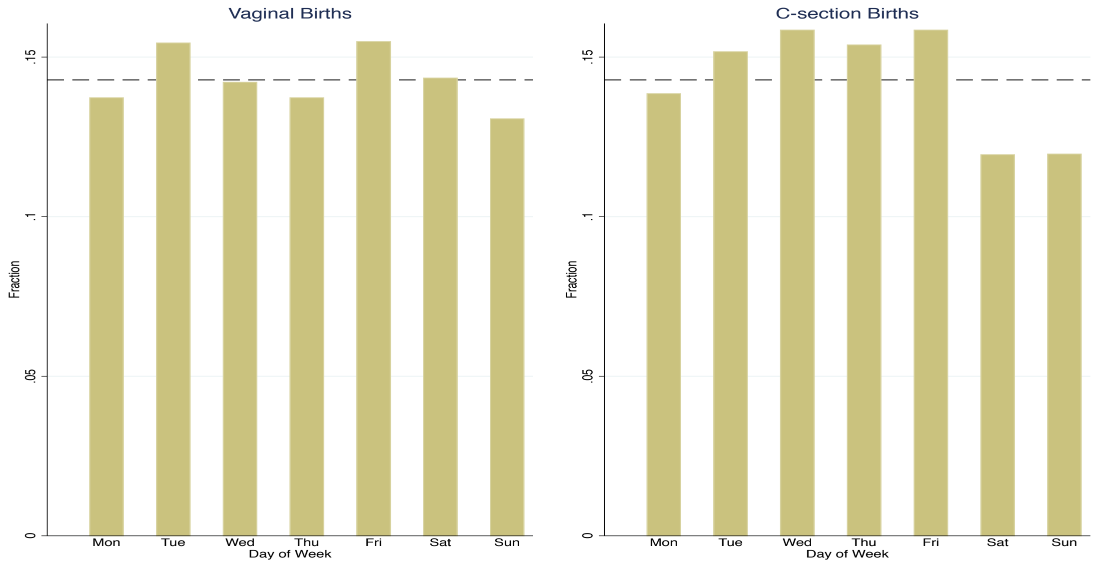
The quasi-randomness of demand shocks in this study arises from the rareness of high-risk admissions and the lack of predictability of vaginal delivery birth time. Very low birth weight newborns are only 1.5% of total births and the median time gap between two vaginally-delivered very low birth weight births in my sample is 14 days. The quartiles of within-hospital median gaps are 13, 22.5, and 40.5 days, indicating that admissions of vaginally-delivered very low birth weight newborns are not common events even in hospitals handling high-risk births relatively frequently. C-section births can either be scheduled or unscheduled (emergent). However, such information is absent in my data. Figure 1 and B1 show time distributions of high-risk newborn admissions by delivery methods. The left panels suggest that vaginal delivered high-risk admissions are evenly distributed in time where we cannot reject the null hypothesis that the fraction of high-risk vaginal births on each day of the week is uniformly distributed. The right panels show decreases in c-section high-risk admissions on weekends and in early mornings. This non-smooth time pattern indicates that at least some c-section high-risk births are scheduled.[2]

To further test for exogeneity, I implement LASSO with cross-validation to predict the occurrence of vaginally-delivered and c-section high-risk newborn admissions. Lagged hospital-day level covariates[3], together with month and day of week indicators, are fed into

[2]There could also exist selection in delivery methods such as births on weekend are less likely to receive c-section. However, if assuming the occurrence of very low birth weight is evenly distributed over time, such hypothesis would predict increases in high-risk vaginal births on weekends and early mornings which is not observed in the data. Furthermore, the attending physicians of newborns are pediatricians and neonatologists, who are expect to have limited coordination with obstetricians. Any endogenous decisions made by obstetricians will not necessarily spillover to the pediatrician side.

[3]Lagged covariates include daily newborn admissions, faction of White/Black/female/low birth weight/very low birth weight/c-section admissions, NICU occupancy, NICU congestion level, number of admitting physicians, average physician experience with low birth weight/very low birth weight patients, and average physician practice style measures in the past three days.

Figure 1: Distribution of High-risk Admissions: Day of Week



Note: The dash line is at $\frac{1}{7}$, showing an hypothetical uniform distribution.

the LASSO prediction model. Hospital-year fixed effects are included but not penalized in the model estimation. None of the lagged covariates show prediction power in the LASSO model for vaginally-delivered high-risk admissions. This provides evidence supporting the hypothesis that the occurrence of vaginally-delivered high-risk admissions is quasi-random. The LASSO model for c-section high-risk admissions, reported in Table A19, indicates that some lagged patient and physician profile measures, as well as certain month and day of week dummies, do show prediction power.

The quasi-random admission time of vaginally-delivered very low birth weight newborns allows me to analyze hospitals' responses to unscheduled high-risk admissions under an event study framework. It also provides a natural experimental setting to study the spillover effects on incumbent newborn patients. Bt assessing the downstream effects on care utilization and patient outcomes among incumbent newborns, I develop a measure of productivity in medical care provision facing fluctuations in demand.

3.2 Hospital Discharge Data

This study utilizes patient admission information from the hospital discharge data collected by the New York State Department of Health (NYSDOH) under the Statewide Planning and Research Cooperative System (SPARCS). All facilities[4] certified to provide inpatient services, ambulatory surgery services, emergency department services or outpatient services are required to submit data to SPARCS. SPARCS collect patient level details on patient characteristics, diagnoses and treatments, services, and charges for each hospital inpatient stay. Patient age is reported in days for patients younger than one year old. The principal diagnosis codes for newborns provide information on multiple births and birth delivery methods. I include all observations with a newborn principal diagnosis code and an age zero day in my sample of birth admissions ("newborn sample" thereafter). Birth weight is reported in grams for all newborn admissions. I assign a low birth weight (LBW) indicator to newborns with birth weight between 1500 and 2500 grams and a very low birth weight (VLBW) indicator to newborns with birth weight below 1500 grams. Patient admission and discharge time (date and hour) reported on the discharge records are used to identify the time of unscheduled high-risk admissions and assign newborns to the treated and control groups (more details described in Section 5.2).

Hospital discharge data provide a unique patient identifier (UPI) that can be used to trace medical records of the same patient across hospitals over time. I construct a panel of hospital admissions within one year after birth for each newborn in my sample. Two binary readmission variables are derived to measure patient outcomes. Neonatal readmission is defined as having readmissions within 28 days after birth. One-year readmission is defined as having readmissions within one year after birth. In addition, one-year cumulative care utilization, i.e. length of stay, total charges, and number of treatment procedures, are aggregated across all hospital admissions during the first year after birth, including one's birth admission. These outcome measures are matched to each newborn's birth admission record. Only the birth admission record for each newborn is included in the analysis sample.

3.3 Physician Characteristics

Each patient admission is assigned with an attending physician on the discharge record. Physician license information from the New York State Education Department (NYSED) Office of Professions is matched to patient admissions by a unique state physician license identifier. The key physician license information used in this study is the date of licensure

[4]Different facility locations under the same operating hospital are regarded as separate facilities in the SPARCS data.

which is used to calculate physician tenure. A very small fraction of attending doctors in the newborn sample are not licensed physicians or were licensed outside of New York State, for whom no license profile can be matched. The missing rate is below 3%.

I allow attending physician characteristics to evolve over time. For each patient admission, physician tenure is computed as the number of years between the year of admission and the year of physician licensure. Since physician specialty is not listed on the license, I define a measure of "experience with high-risk newborns" as a proxy for physician expertise. For patients admitted on day t attended by physician p, the physician expertise is measured by the fraction of newborn patients being VLBW among all newborn patients attended by physician p up to day $t-1$. In addition, I develop physician practice style baseline measures using average total charges, length of stay, and number of procedures among newborn patients *discharged* up to day $t-1$ by physician p. The averages are only measured up to the day before one's admission, hence eliminating any influence from the patient's own admission or future admissions. These baseline measures allow physicians' practice styles to evolve with the patient conditions they treated in the past.[5] To account for patient–physician selection, residual total charges, length of stay, and number of procedures are generated controlling for hospital-year fixed effects and patient observables. These residual measures capture physician practice style conditional on patient observables and hospital-year specific effects, hence may be interpreted as physician "intrinsic" styles. However, due to the presence of patient–physician selection, the effects of patient observables cannot be well identified. This will result in biased physician residual measures.[6] Therefore, I take raw averages as the preferred physician measures in this study.

3.4 NICU Daily Census and Congestion

I construct a NICU daily patient census to measure the level of congestion. The UB-04 revenue codes on hospital discharge records provide information on the type of accommodation and the number of days of each accommodation one received during the hospital stay. I follow Freedman (2016) and flag revenue codes of 1703 - Nursery Level III ("Intermediate Care") and 1704 - Nursery Level IV ("Intensive Care") as NICU accommodations.

[5] A physician with high average in total charges could be due to spending more on an average patient or treating more high-spending patients.

[6] To illustrate, if physicians with intensive practice styles always treat high-risk patients, then the effect of practice style cannot be separated from the effect of patient condition.

The accommodation types are listed in chronological order, which allows me to derive the admission and discharge dates of each NICU patient. The daily NICU patient census is derived based on the universe of patient admissions, regardless of whether they are in my newborn sample. Using each patient's NICU admission and discharge dates, I derive the number of NICU admissions, NICU discharges, and NICU patient occupancy on each hospital-day. Since the hospital discharge data in 2005 include patients who were admitted in 2004 and discharged in 2005, the NICU daily occupancy measure is precise from the first day of 2005.

Figure B3 shows the distribution of daily NICU occupancy in terms of a fraction of the annual median occupancy in the same NICU (i.e. the variation reflects within hospital fluctuations). Daily NICU congestion level is defined based on the quartiles of daily NICU occupancy within each hospital-year. The top quartile hospital-days are coded as high congestion level, the bottom quartile as low congestion level, and the middle two quartiles as medium congestion level. This occupancy measure better captures the level of *relative* congestion compared to using the daily number of empty NICU beds due to the following reason: hospitals can frequently keep "temporary" NICU beds which usually are not shown in official hospital facility reports. Hence, the actual capacity can go beyond the officially reported bed capacity. This pattern is recognized by NYSDOH[7] and empirically observed in the hospital discharge data. I follow Freedman (2016) to obtain NICU bed counts from hospital annual Institutional Cost Report (ICR). I also count NICU beds registered under the Certificate of Need (CON) system to cross validate the NICU bed capacity. The NICU bed counts reported by hospitals in their annual ICR differ from their CON registered NICU bed counts. In addition, the daily NICU occupancy derived from the hospital discharge data exceeds the reported NICU bed capacity on a frequent basis. This implies that hospitals frequently adopt "temporary" NICU beds to expand their NICU capacity. In this case, using a relative NICU congestion measure better identifies periods when hospitals indeed face resource constraints relative to their normal patient volume accounting for any "temporary" beds they may use.

[7] NYSDOH issued a letter in 2016 noting that "*It has come to the New York State Department of Health (Department)'s attention that bed capacity in New York State neonatal intensive care units (NICU) is being exceeded on a frequent basis.*" https://www.health.ny.gov/professionals/hospital_administrator/letters/2016/2016-09-27_dal_16-14_nicu_vercrowding.htm

3.5 Sample Description

This study focuses on newborn birth admissions to hospital facilities in the New York City area[8] during 2005 to 2009. There are 46 hospitals in the New York City area with live birth admissions during the sample period. This study is limited to 36 hospitals with non-zero annual NICU accommodations and further exclude 2 hospitals with fewer than 100 annual birth admissions. The resulting analysis sample consists of 489,635 newborn birth admissions in 34 New York City area hospitals with NICU facilities.

Birth weight provides a good metric of newborn health and expected care utilization. Figure B4 and B5 show the average total charges and the in-hospital mortality by birth weight categories in the sample. VLBW newborns are only 1.5% of total births but demand 33% of total newborn care medical spending and have average in-hospital mortality as high as 16.6%. LBW newborns constitute 7% of total births, consume 22% of newborn care spending, and have higher in-hospital mortality compared to normal birth weight newborns. I define newborns with birth weight below 1500 grams, i.e. VLBW newborns, as high-risk admissions, newborns with birth weight between 1500 to 2500 grams, i.e. LBW newborns, as mid-risk admissions, and newborns with birth weight of 2500 grams and above as low-risk admissions.

The newborn sample consists of 7,448 high-risk admissions (2,273 vaginally-delivered and 5,175 c-section), 33,158 total mid-risk admissions (17,105 vaginally-delivered and 16,053 c-section), and 449,029 total low-risk admissions (309,637 vaginally-delivered and 139,392 c-section). The birth weight distribution in the sample is shown in Figure B6. Figure B7 - B10 depict the distribution of attending physician characteristics over patient birth weight. Newborns with lower birth weight are treated by physicians with longer tenure, more experience with high-risk newborns, and higher care utilization. Attending physicians of healthier newborns are more junior, less experienced with high-risk newborns, and have less intensive practice styles. In addition, there is a clear decreasing trend in each attending physician measure among mid-risk newborns when birth weight increases from 1500 grams to 2500 grams.

Table 1 reports summary statistics in the full newborn sample and in subsamples by newborn birth weight categories. Care utilization increases dramatically with lower birth weight. Low-risk newborns on average stay in hospital for 2.8 days (median 2 days), in-

[8] New York County, Bronx County, Kings County, Queens County, and Richmond County

Table 1: Summary Statistics

Sample[a]	(1) All	(2) Low-risk	(3) Mid-risk	(4) High-risk
Panel A: Care Utilization and Patient Outcomes				
Length of Stay	3.917	2.791	8.887	49.709
(Median)[b]	2	2	4	45
Total Charges	12,272	6,050	40,118	263,435
(Median)[b]	3,813	3,638	11,940	180,016
Number of Procedures	1.693	1.528	2.706	7.114
NICU Admission	0.158	0.115	0.564	0.962
Death in Hospital	0.003	0.000	0.006	0.166
Hospital Transfer	0.004	0.002	0.013	0.091
28-Day Readmission	0.013	0.013	0.016	0.008
1-Year Readmission	0.043	0.040	0.068	0.141
Panel B: Attending Physician Characteristics				
Physician Tenure	16.841	16.811	17.118	17.405
Physician Experience with VLBW	0.017	0.013	0.050	0.101
Physician Experience with LBW	0.072	0.065	0.136	0.204
Physician Average Length of Stay	3.796	3.512	6.314	9.773
(Median)[b]	2.709	2.668	4.134	7.968
Physician Average Total Charges	10,937	9,179	26,455	48,195
(Median)[b]	4,990	4,869	7,549	27,723
Physician Average Number of Procedures	1.582	1.521	2.132	2.808
Observations	489,635	449,029	33,158	7,448

[a] Low-risk sample consists of newborns with birth weight of 2500g and above. Mid-risk sample consists of newborns with birth weight between [1500, 2500)g, i.e. LBW. High-risk sample consists of newborns with birth weight below 1500g, i.e. VLBW.
[b] Median is reported for length of stay and total charges only because the distribution is heavily skewed. Mean and median are close for other continuous variables.

cur total charges of $6.1k (median $3.6k), and have a NICU admission rate of 11.5% after birth. Mid-risk newborns on average stay in hospital for 8.9 days (median 4 days), incur total charges of $40.1k (median $11.9k), and have a NICU admission rate of 56.4%. High-risk newborns demand intensive care after birth. They stay in hospital for 49.7 days (median 45 days), incur total charges of $263.4k (median $180.0k), and have a NICU admission rate of 96.2%. Despite the highly skewed care utilization distribution, newborns with lower birth weight still have worse health conditions upon discharge. In-hospital mortality is extremely low among low-risk newborns, but rises to 0.6% among mid-risk and 16.6% among high-risk newborns. The average one-year readmission rate among low-risk newborns is 4%. This rate is 6.8% for mid-risk newborns and 14.1% for high-risk newborns. High-risk newborns have a low 28-day readmission rate because they are likely not yet discharged on the 28th day. Table 1 Panel B reports average attending physician characteristics. Physician tenure and experience with high-risk newborns increase with decreased birth weight. Attending physicians of lower birth weight newborns also have more intensive practice styles, shown by longer length of stay, higher total charges, and higher number of procedures.

4 Hospital Response to Unscheduled High-risk Admissions

Admissions of vaginally-delivered VLBW newborns are defined as unscheduled high-risk admissions in this study and serve as demand shocks to hospitals. Section 3.1 and 3.5 provide empirical evidence that unscheduled high-risk admissions happen quasi-randomly in time and result in high care utilization. In this section, I answer the question of how hospitals respond to unscheduled high-risk newborn admissions and explore whether the response varies by the level of congestion, measured by NICU patient occupancy. I find that hospital capacity can be adjusted in certain extent to accommodate short-term fluctuations in care demand. Instead of competing for or crowding out existing resources allocated to previously admitted newborns, unscheduled high-risk admissions draw in additional resources such as highly specialized physicians. This spills over to incumbent newborn patients that the highly specialized physicians also treat previously admitted newborns, especially when hospitals are congested.

4.1 Event Study Specification

I study hospitals' responses under an event study framework. Patient level data are aggregated into hospital-day or hospital-physician-day panels. Any hospital-day or hospital-physician-day with vaginally-delivered VLBW newborn admissions are defined as an event and labeled as "day 0". The 2,273 vaginally-delivered VLBW newborn admissions in the sample constitute 2,156 events.[9]

The event study is implemented in a 5-day window centered at the day of unscheduled high-risk admissions. In the case of overlapping event windows, multiple event day

[9]Multiple VLBW newborns can be admitted to the same hospital on he same day, such as twin births.

indicators are assigned to the same hospital-day.[10]

$$Y_{h,(p),t} = \sum_{j=-5}^{5} \phi^j D^j_{h,(p),t} + (\alpha_p) + \tau_{h,y} + \tau_{dow} + \tau_m + \epsilon_{h,(p),t} \tag{1}$$

- $Y_{h,(p),t}$ is the outcome of an admission in hospital h on day t (by physician p).
- $D^j_{h,(p),t}$ are event time indicators: $D^j_{h,(p),t} = 1$ for being j days apart from an event.[11]
- $D^{-1}_{h,t}$ is excluded as the baseline event period and ϕ^{-1} is normalized to zero.
- $\tau_{h,y}$, τ_m, τ_{dow} are fixed effects for admission hospital-year, month, and day of week.
- α_p is physician fixed effect, included in analyses of hospital-physician-day panels.

The event study coefficient ϕ^j measures hospitals' responses to an event j days from the day of event. ϕ^1 is normalized to zero with $D^1_{h,(p),t}$ omitted in the regression. Coefficient on the day of event, ϕ^0, captures any spontaneous responses to the unscheduled high-risk admission. Post-event coefficients ϕ^j, $j > 0$, measure any lasting effects or delayed adjustments. Pre-event coefficients ϕ^j, $j < 0$, provide a test on exogeneity: any pre-event responses would suggest that the subsequent high-risk newborn admissions may be expected or the high-risk admission decisions are made endogenously.

4.2 Graphical Evidence of Research Design

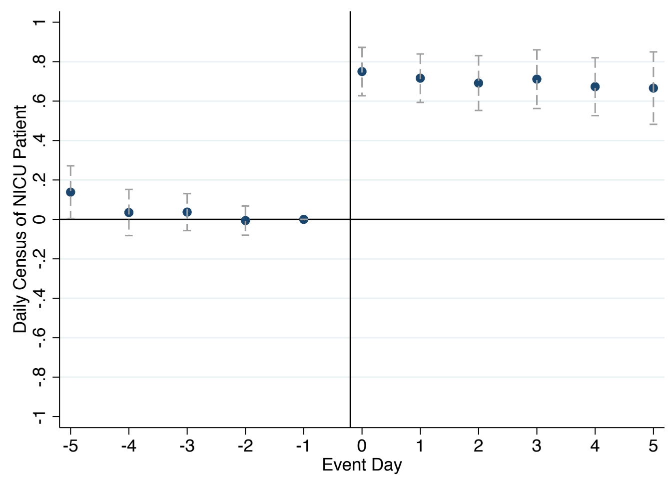
I start by presenting event study figures on patient flow, showing how admissions of unscheduled high-risk newborns affect daily NICU occupancy, admissions, and discharges. Figure 2 plots event study coefficients on daily NICU occupancy. It shows that the number of NICU patients increases by 0.75 on the day of event and that this increase persists throughout the 5 subsequent days. Such an increase indicates that the unscheduled high-risk admissions lead to a sharp and persistent short-run increase in care demand. Figure B11 demonstrates the time pattern of NICU admissions and discharges in the 5-day event study window. The increase in NICU occupancy on the day of event is driven by an average increase of 0.91 in NICU admission and an average increase of 0.16 in NICU discharge. The magnitude of the increase in admission is consistent with the high NICU admission

[10] Consider two events 3 days away in the same hospital. In this case, the hospital-day before the first event will be assigned with event day indicators -1 and -4. 39.5% of the event study sample is assigned with multiple event day indicators.

[11] $D^j_{h,(p),t} = 1$ if $Y_{h,(p),t}$ is $|j|$ days before an unscheduled high-risk admission in hospital h (treated by physician h) for $j < 0$ and j days after the unscheduled high-risk admission for $j > 0$.

rate for VLBW newborns: more than 90% of the high-risk newborns are directly admitted to NICU upon birth.[12] The increase in discharges on the day of event is entirely driven by same-day discharges of the unscheduled high-risk newborns who are either transferred out from NICU or do not survive.

Figure 2: Daily Census of NICU Patients



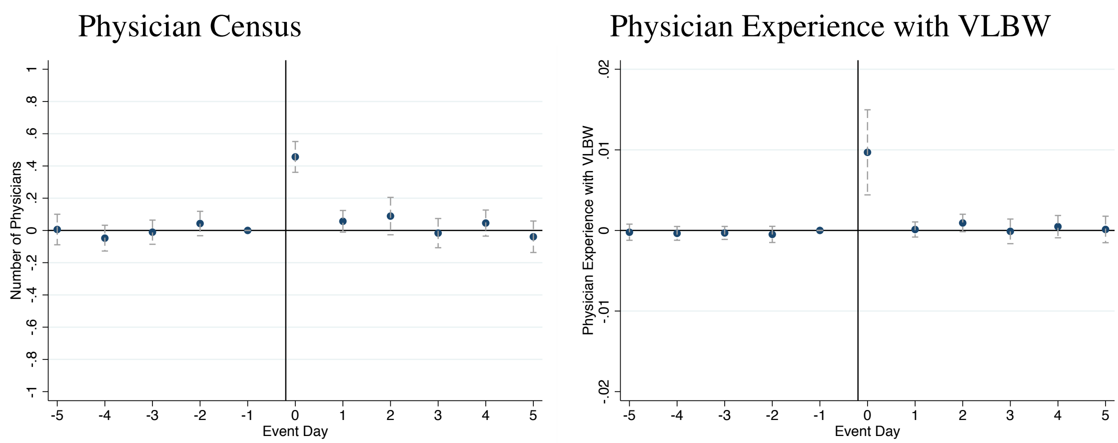
Existing studies show that hospitals may discharge patients early or reduce subsequent admissions to manage fluctuations in patient flow. Since the high-risk newborn admissions increase both patient volume and average acuity in NICU, we might expect hospitals to adopt similar managing strategies, especially when NICU occupancy is high. However, the post-event coefficients in figure B11 show no such patterns. I further explore the effect of NICU congestion by focusing on events when the NICU facilities are congested on the day before unscheduled high-risk admissions. Analyses conditioning on NICU congestion levels will mechanically induce mean reversion patterns in patient flow.[13] To account for this, I construct a control group by randomly sampling hospital-days with the same congestion restriction as "placebo" events. Any mean reversion patterns are differenced out and DD-event study coefficients are plotted in B12. The DD-event study coefficients show neither increases in NICU discharges nor decreases in NICU admissions during post-event periods, suggesting that hospitals do not face hard capacity constraints and attain some flexibility in accommodating short-term demand shocks even when the occupancy level is

[12]High-risk newborns not admitted to NICU are either transfered to a different medical facility or too sick to receive any NICU care.

[13]The mean reversion pattern will show an increase (decrease) in admission (discharge) prior to the event and a decrease (increase) in admission (discharge) post to the event if we require NICU to be congested upon the event.

high. The event study figure on daily NICU census, combined with the empirical observation that NICU occupancy can frequently exceed their registered bed capacities, implies that hospitals can increase facility resources by adding "temporary" beds to accommodate fluctuations in patient flow.

Figure 3: Effects on in Physician Census and Profile



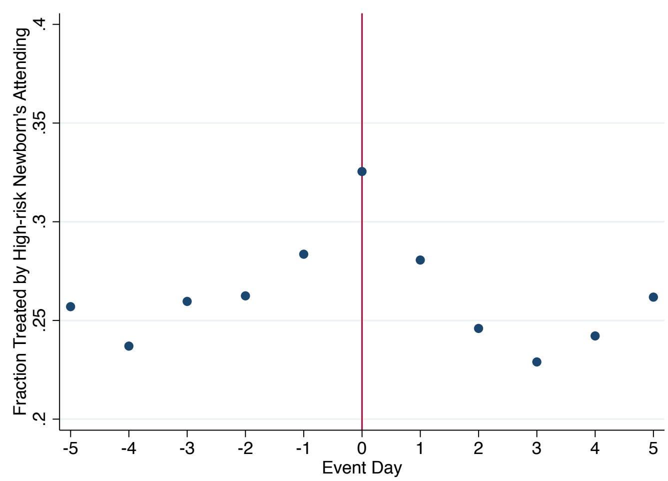
Switching from facility resources to physicians, I investigate how hospitals manage physicians in response to unscheduled high-risk admissions. Figure 3 left panel points to an average increase of 0.5 attending physicians when unscheduled high-risk newborns are admitted. The increase in newborn patient volume is fully offset by the additional physicians on the day of event, keeping the patient–physician ratio unaffected (Figure B13). Figure 3 right panel presents the change in physician composition: the attending physicians on the day of event are more experienced in treating high-risk newborns, which reflects hospitals' responses to the increase in patient acuity.[14] The findings on physician census and physician composition suggest that hospitals bring in additional physicians specialized in treating high-risk newborns to accommodate the unexpected high-risk admissions.

The event study results indicate that hospital capacity can be adjusted to a certain extent to accommodate short-term fluctuations in care demand. Instead of competing for or crowding out existing resources allocated to previously admitted newborns, unscheduled high-risk admissions draw in additional resources such as highly specialized physicians. The increase in available resources will not only be allocated to high-risk newborns which usually produces health benefits (Almond et al., 2010; Chyn et al., 2019), but it may affect

[14]Figure B8 plots the distribution of physician experience in treating high-risk newborns over birth weight and indicates that high-risk newborns are treated by physicians specialized in handling high-risk cases.

incumbent newborns as well. To explore any potential spillovers, Figure 4 plots the fraction of mid-risk newborn admissions attended by the attending physicians of unscheduled high-risk newborns ("specialized physicians") on each day in the 5-day window. 32.5% of mid-risk newborn admissions admitted on the same day as the unscheduled high-risk newborns are assigned to the specialized physicians. This fraction drops to 28% on the day before or after the event.[15] Figure 5 presents the same figures conditional on NICU congestion levels on the day before unscheduled high-risk admissions (i.e. event day -1). When NICUs are congested, the specialized physicians treat a higher fraction of previously admitted mid-risk newborns. On the contrary, such fraction is lower if NICU facilities are not congested upon unscheduled high-risk admissions and the specialized physicians are more likely to treat subsequent mid-risk newborn admissions. The changes in physician–patient assignment among incumbent newborns implies further spillover effects on care utilization and patient outcomes. These effects, with their implications for medical care productivity, are presented in Section 5.

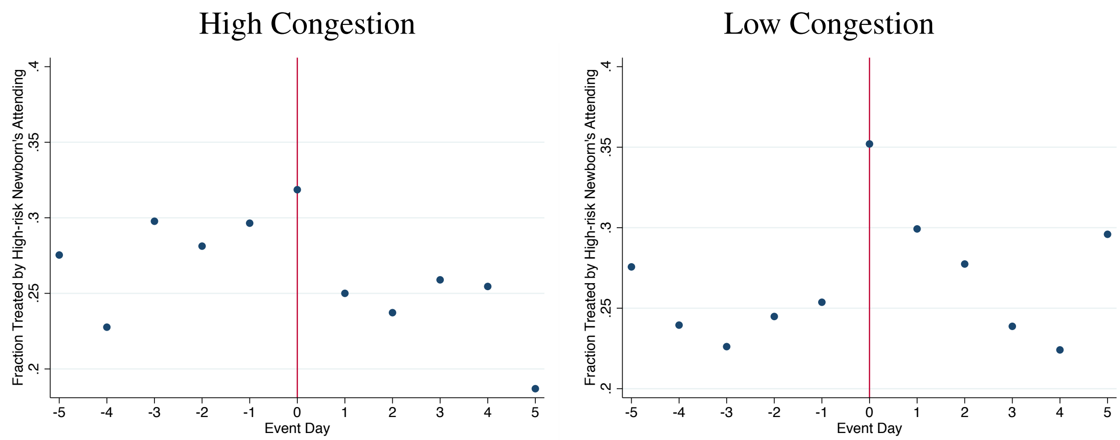
Figure 4: Fraction of Mid-risk Newborns Treated by Specialized Physicians



The hypothesis that hospitals call in specialized physicians upon unscheduled high-risk admissions which spills over to incumbent patients is well supported by the event study findings. However, a data limitation prevents validating it with any direct evidence. In the hospital discharge data, a physician can be observed only when being assigned as an

[15] A patient's attending physician can be different from the admitting physician. The attending physician is the one making most treatment decisions and can be assigned later during the stay. It is not unusual that a physician gets in hospital and receives a patient admitted on the previous day.

Figure 5: Fraction of Mid-risk Newborns Treated by Specialized Physicians



attending physician. The time associated with physician presence is the admission time of the patient, not the time when the physician actually treats the patient or starts working. In the case that a specialized physician is in the hospital treating patients during pre-event days and then gets assigned to the unscheduled high-risk patient on the day of event, it will generate the same pattern as in Figure 4. Hence, this data limitation makes it difficult to distinguish between physicians who are called in and physicians who have been on duty at the event time. However, whether the specialized physicians are called in or have been on duty, the research setting in this study stays valid as long as unscheduled high-risk admissions create exogenous variation in physician–patient matching among previously admitted newborns not associated with patient conditions.

To provide supporting evidence that hospitals call in specialized physicians upon unscheduled high-risk admissions, Figure B14 overlays two event study plots on daily admission census for the specialized physicians at a hospital-physician-day level.[16] Comparing events when a specialized physician treats the unscheduled high-risk newborns (high-risk events) to events when the physician attends other newborn admissions (placebo events)[17], these physicians treat fewer patients on days prior to and following treating an unscheduled high-risk newborn compared to their usual admitting pattern. This indicates that some physicians may be called in just upon unscheduled high-risk newborn admissions. Figure B15 plots event study coefficients for high congestion events. When the NICU facilities are

[16]The number of patients attended includes both newborn patients and non-newborn patients. On average, 86% of the patients treated by the specialized physicians are newborn patients, and the average fraction of under 5 years old is 95%.

[17]Placebo event dates are constructed under stratified random sampling which preserves the number of events in each physician-hospital cell.

congested, gaps on event day -2 and -3 are wider, the gap on event day -1 is slightly smaller, and physicians attend more patients in general. This pattern is consistent with the hypothesis that hospitals are more likely to call in specialized physicians when NICU facilities are congested (larger gaps on event day -2 and -3) and these physicians treat more patients admitted on the previous day (smaller gap on event day -1). Figure B16 provides supporting evidence that physicians on average attend fewer patients on days prior to attending an unscheduled high-risk admission.

5 Spillover Effect of Unscheduled High-risk Admissions

Event study results in Section 4 indicate that unscheduled high-risk admissions can affect attending physician assignment among previously admitted newborns, especially when the NICU congestion level is high. This exogenous variation in physician–patient matching provides a natural experimental setting to study the effect of physician practice on patient outcomes. In this section, I implement patient level regressions and report estimated spillover effects on physician practice styles, care utilization, and patient outcomes. The results indicate that incumbent newborns received more intensive treatments but show little health improvement. The increase in care utilization is likely driven by physician practice styles. I also present a stylized production model in Appendix Section A to rationalize hospitals' responses and discuss the implications for efficiency.

5.1 Spillover Effects Predicted by Hospital Response

Event study findings provide empirical evidence that hospitals respond to unexpected increases in demand differently based on the level of congestion. The stylized hospital production model in Appendix Section A helps rationalize the differential responses by showing that the optimal matching decision depends on the relative magnitudes of a) the productivity gains of matching physicians with suited skills to patients; and b) the incurred costs of achieving good skill–task matching between physicians and patients.

Figure B17 lists four hypothetical hospitals' responses. Assuming physician A, B, and C are on site upon the admission of an unscheduled high-risk patient. Since the potential cost of mistreatment is consequential on high-risk patients, hospitals are likely to prioritize the unscheduled high-risk admissions when making physician assignments. According to the empirical patterns in Figure B7 - B10, we assume that the attending physician of the high-risk patient has a more intensive practice style and is more experienced in treating high-risk patients. Scenario 1 and 2 hypothesize that the high-risk patient is assigned to an on-site physician and no new physician is called in. In scenario 3 and 4, a highly specialized physician D is called in to treat the high-risk patient.

Whether hospitals call in any specialized physicians and whether specialized physicians treat any incumbent patients lead to different predictions of spillover effects. Scenario 1 predicts that physicians treating incumbent patients will be less experienced in treating high-risk patients and have less intensive styles. Scenario 4 predicts that incumbent patients will be assigned to physicians with more intensive styles and more experienced in treating high-risk patients. Scenario 2 and 3 predict relatively little or no spillover. If we believe that hospitals are in higher need for additional physicians when they are congested and that the called in physicians are also more likely to treat incumbent patients as shown in Figure 5 left panel, we would expect to observe increased average level of physician specialization and style intensity among incumbent newborns when NICU occupancy is high, as suggested in Scenario 4. If we believe that hospitals are less likely to call in additional physicians or that specialized physicians are less likely to treat incumbent patients as shown in Figure 5 right panel, we would expect to observe reduced average level of physician specialization and style intensity among incumbent newborns when NICU occupancy is low, as suggested in Scenario 1.

5.2 Regression Specification

To estimate the spillover effects of high-risk admissions on previously admitted newborn patients, I partition the newborn sample into four subsamples, listed in Table A1. Following the definition in event study, vaginally-delivered VLBW newborn admissions are defined as unexpected high-risk admissions. C-section VLBW newborn admissions are excluded in the spillover analyses, because they are not smoothly distributed over time as shown in Figure 1 and B1.[18] I focus on mid-risk newborns in measuring the spillover effects for two reasons: 1) mid-risk newborns demand medical care upon birth, and are hence vulnerable to demand shocks; 2) mid-risk newborns have higher rate of NICU admission, and are hence more likely to share medical resources with high-risk newborns in

[18] I provide supplementary analyses using C-section VLBW newborn admissions as unscheduled high-risk admissions in Appendix Section B

the NICU. Low-risk newborns mostly stay in the regular nursery after birth and require little medical care. They serve as a placebo group in this study since unscheduled high-risk admissions are expected to have little impact on these healthy newborns. All multiple births are excluded from the spillover analysis sample.

The mid-risk analysis sample consists of 23,791 newborn admissions with birth weight between 1500 grams to 2500 grams. Newborns admitted within two days prior to an unscheduled high-risk admission are assigned to the treated group. The control group consists of all newborns whose birth admission lies three or more days apart from any unscheduled high-risk admissions. Newborns admitted on the same day or within the two days after an unscheduled high-risk admission are studied separately in supplementary analyses in Appendix Section B since these newborns may experience very different spillover effects compared to incumbent newborns. Their admissions can also be endogenous to the unscheduled high-risk admissions.[19]

Adopting an admission time cutoff in assigning newborns to treated and control groups is essential to this study. Using actual overlaps with unscheduled high-risk newborns would lead to bias or fail to capture key effects for several reasons: 1) length of day is an outcome which could be affected by the unscheduled high-risk admission hence is endogenous; 2) newborns with longer length of stay tend to have worse health conditions and have higher probability of encountering unscheduled high-risk admissions; 3) mid-risk newborns might have reduced demand for care after several initial days, and are therefore expected to show little effect if encountering unscheduled high-risk admissions too late during their hospital stay. The admission time cutoff is chosen empirically to balance the actual rate of overlap and care intensity demanded. Less than 2% of low birth weight newborns are discharged on the day of birth or on the day after (Figure B18). Therefore, birth admissions within the two days prior to an unscheduled high-risk admission almost surely have some overlap with the high-risk newborn, shown in Figure 6. In addition, as shown in Figure B19, care intensity is concentrated in the first three days during newborns' hospital stays. To establish direct evidence that unscheduled high-risk admissions lead to sharp increases in care demand, Figure 7 counts the total number of procedures performed on NICU patients and plots the fraction performed on the unscheduled high-risk patients. On the day of unscheduled high-risk admission, the newly admitted high-risk patients take up more than 50% of total NICU

[19]Hospitals might be selective in admitting newborns after unscheduled high-risk admissions which will result in endogeneity. Newborns admitted on the same day of and shortly after any unscheduled high-risk admissions are exposed to such potential selection issue.

procedures. This fraction decreases to approximately 10% on the 3rd day after birth and further to below 5% on the 7th day and thereafter.

Figure 6: Identifying a Treated Group by Overlap with High-risk Admission

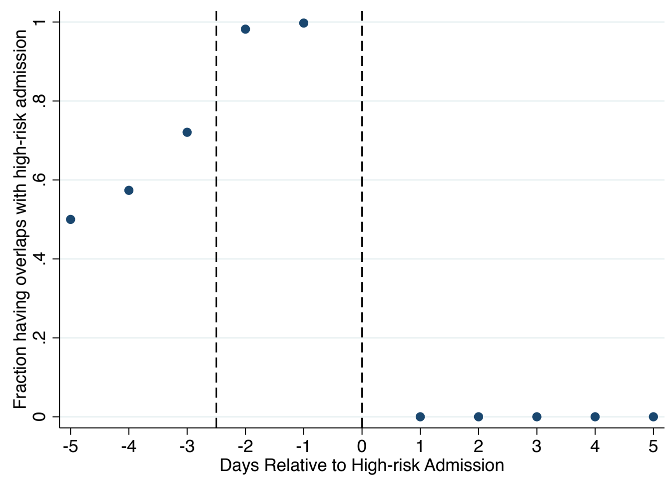
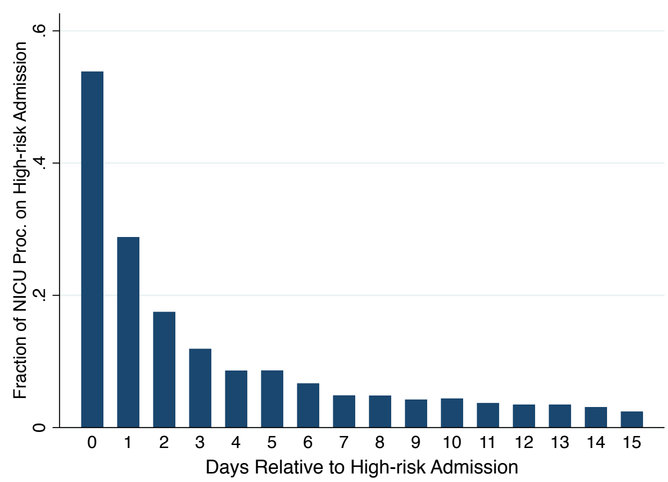


Figure 7: Fraction of Procedures on High-risk Admissions



The key identifying assumption is that encountering an unscheduled high-risk admission within two days after birth is as good as randomly assigned. Under such assumption, newborns in the treated and control groups should be similar upon admission. Any differences in care utilization or patient outcomes would serve as a measure of the spillover effects of unscheduled high-risk admissions. A patient level regression is specified to capture any differences between the treated and control groups:

$$Y_{i,h,t} = \alpha \cdot Pre_{i,h,t} + \beta X_{i,h,t} + \tau_{h,y} + \tau_{dow} + \tau_m + \epsilon_{i,h,t} \tag{2}$$

- $Y_{i,h,t}$ is the outcome measure of newborn i admitted on day t in hospital h.
- $Pre_{i,h,t}$ is the treated group indicator: $Pre_{i,h,t} = 1$ for newborns admitted within the 2 days prior to an unscheduled high-risk admission.
- $X_{i,h,t}$ flexibly controls for patient observables, including dummies for birth delivery method, insurance type, race, gender, and birth weight (250-gram bins).
- τ_{hy}, τ_m, τ_{dow} are hospital-year, birth month, and day of week fixed effects.

α is the coefficient of interest in this study, which captures any differences between the treated and control groups. It provides a measure of the spillover effect if the outcome variable is care utilization or patient outcome measures. When substituting patient observables as outcome variables in the regression, coefficient α provides a direct test of the identifying assumption.

To examine the effect heterogeneity across congestion levels, I interact the treated group indicator with the congestion indicator following the following regression specification:

$$Y_{i,h,t} = \sum_c \alpha_c \cdot Pre_{i,h,t} \times (Cgst_{h,t} = c) + \beta X_{i,h,t} + \sum_c \gamma_c + \tau_{h,y} + \tau_{dow} + \tau_m + \epsilon_{i,h,t} \tag{3}$$

- $Cgst_{h,t}$ is the NICU congestion indicator described in Section 3.4.[20]
- α_c captures spillover effects at each congestion level.
- γ_c controls for base level differences by congestion levels.[21]

To allow flexibility, I implement an augmented regression model by interacting the congestion indicator with all covariates and fixed effects. The augmented regression model is equivalent to subsample regression by congestion levels following equation (2). Coefficient estimates without covariates-congestion interactions from regression (3) and regression estimates from the augmented regression model are both reported for comparison.

Table 2 presents covariates balance in the sample, estimated under equation (2). Patient

[20]The congestion level is measured on the day before each newborn's admission. The congestion level upon newborn admissions in the treated group persists till the day of the unscheduled high-risk admission.

[21]Congestion level could affect hospital admission decisions. Patients admitted when hospitals are congested can differ from patients admitted when hospitals are not congested. Hence, it is essential to match the congestion level for the control and treated groups so that the treated–control differences are causally interpretable.

observables X are removed from the regressors and used as dependent variables. Columns 1 and 2 report covariate averages in the control and treated groups. Columns 3 - 6 report average covariate differences between the treated and control groups, controlling for hospital-year, month, and day of week fixed effects. We observe small and insignificant differences between the treated and control groups across all measures in the overall sample and at each congestion level. This provides strong support that the treated–control status is as good as randomly assigned.

Table 2: Covariates Balance Table

	(1) Control	(2) Treated	(3) Difference	(4) Difference	(5) Difference	(6) Difference
C-section	0.409	0.392	-0.0152	-0.0137	-0.0137	-0.0270
			(0.0123)	(0.0186)	(0.0198)	(0.0223)
Birth Weight	2197.2	2184.0	-5.043	-4.055	-0.625	-16.24
			(6.373)	(12.27)	(10.03)	(13.72)
White	0.258	0.295	-0.00481	-0.00684	-0.00138	-0.0229
			(0.00957)	(0.0148)	(0.0129)	(0.0185)
Black	0.333	0.321	-0.00363	-0.00826	-0.00342	0.00152
			(0.0115)	(0.0233)	(0.0132)	(0.0273)
Female	0.534	0.528	-0.00888	-0.0167	-0.0189	0.0270
			(0.00822)	(0.0187)	(0.0130)	(0.0261)
Medicaid	0.611	0.599	0.00387	-0.00344	0.00360	0.0135
			(0.00760)	(0.0166)	(0.0117)	(0.0191)
Observations	21629	2162	23791	7015	11790	4986
Congestion	All	All	All	Low	Medium	High

Standard errors in parentheses. Standard errors are clustered at hospital level

[a] Hospital-year, month, and day of week fixed effects are included in measuring treated–control differences.

5.3 Impact Estimates

In this section, I report the estimates of spillover effects from regression equations (2) and (3). Two sets of outcomes are analyzed: 1) attending physician profile and practice style; 2) care utilization and patient outcomes. In Table 3 - 6, regression coefficients from equation (2) are reported in the top panel and regression coefficients at each congestion level are reported in the bottom panel. The effect estimates attain similar magnitudes with or without congestion-covariates interactions. Hence, I only report estimates with congestion-covariates interactions from the subsample regression discussed in Section 5.2

in alternative specifications and robustness checks.

5.3.1 Attending Physician Profile and Practice Styles

As motivated in Section 5.1, newborns admitted prior to unexpected high-risk admissions may experience either positive or negative spillovers in terms of physician–patient matching. This section summarizes the spillover effect on attending physician tenure, experience with sick newborns, and physician practice style measures.

Table 3 reports differences in attending physician profiles. Coefficients in the top panel indicate no overall difference between the treated and control groups. The bottom panel suggests that when NICUs are congested, newborns admitted within the two days prior to unscheduled high-risk admissions are attended by physicians more experienced in treating sick newborns. At high congestion levels, columns 3 - 6 indicate that treated group newborns are assigned to physicians with 6% more experience with high-risk newborns and 10% and more experience with mid-risk newborns. When the NICU occupancy is low, columns 1 and 2 indicate that newborns admitted before unscheduled high-risk admissions are treated by more junior physicians, shown by a decrease in tenure of more than 0.8 years. These physicians are also marginally less specialized in treating sick newborns.

Table 3: Differences in Attending Physician Profile

	(1)	(2)	(3)	(4)	(5)	(6)
Congestion	Physician Tenure		Experience with VLBW^a		Experience with LBW^a	
All	-0.234		0.00138		-0.0000981	
	(0.258)		(0.00146)		(0.00186)	
Low	-0.882**	-0.824*	-0.000221	-0.000908	-0.00448	-0.00619*
	(0.380)	(0.408)	(0.00323)	(0.00317)	(0.00338)	(0.00340)
Medium	-0.114	-0.0789	0.000517	0.000225	-0.00161	-0.00177
	(0.316)	(0.352)	(0.00279)	(0.00295)	(0.00221)	(0.00221)
High	0.406	0.423	0.00594*	0.00606*	0.0101**	0.0117***
	(0.427)	(0.514)	(0.00335)	(0.00314)	(0.00418)	(0.00405)
Covariate x Cgst	No	Yes	No	Yes	No	Yes
N^b	23164	23164	23669	23669	23669	23669
Y-mean	17.13	17.13	0.0459	0.0459	0.130	0.130

Standard errors in parentheses Standard errors clustered at hospital level

[a] VLBW newborns=high-risk newborns; LBW newborns=mid-risk newborns.

[b] A small fraction of patients have no attending physician measures because of missing physician license information or no previous admitted patients to construct experience measure.

Table 4: Differences in Attending Physician Practice Style[a]

	(1)	(2)	(3)	(4)	(5)	(6)
Congestion	Avg. Length of Stay (log)		Avg. Total Charges (log)		Avg. # of Procedures	
All	0.000803		0.00638		0.0326**	
	(0.00736)		(0.0144)		(0.0157)	
Low	-0.0172	-0.0277*	-0.0304	-0.0502	0.0206	0.00142
	(0.0170)	(0.0161)	(0.0324)	(0.0304)	(0.0283)	(0.0282)
Medium	-0.00335	-0.00342	0.00628	0.00531	0.0150	0.0140
	(0.00934)	(0.00952)	(0.0202)	(0.0209)	(0.0202)	(0.0210)
High	0.0364*	0.0409**	0.0597	0.0708**	0.0889**	0.0849**
	(0.0203)	(0.0188)	(0.0377)	(0.0346)	(0.0373)	(0.0348)
Covariate x Cgst	No	Yes	No	Yes	No	Yes
N[b]	23535	23535	23535	23535	23535	23535
Y-mean	1.771	1.771	9.323	9.323	2.085	2.085

Standard errors in parentheses — Standard errors clustered at hospital level

[a] For patients admitted on day t attended by physician p, physician practice style measures are defined to be average total charges, length of stay, and number of procedures among newborn patients discharged up to day t-1 by physician p.

[b] A small fraction of patients have missing physician practice measures because there is no previously discharged patients by their attending physicians.

Table 4 reports differences in attending physician practice style baseline measures. The top panel indicates an overall more intensive style shown by higher average number of procedures. Coefficients in the bottom panel indicate that this effect is driven by the effect at high congestion levels. When NICU facilities are congested, newborns admitted just prior to unscheduled high-risk admissions are treated by physicians who, on average, assigning 4% longer length of stay, 7% higher charges, and 4% (0.0849/2.085) more treatment procedures. Table A2 reports estimated differences in attending physician practice style residual measures. Despite the measurement issue discussed in Section 3.3, point estimates show a similar pattern when compared to Table 4.

Linking the empirical estimates on physician measures to the predictions presented in Section 5.1, Table 3 and 4 indicate that the differences in attending physician profile when NICU congestion level is low are consistent with the prediction in scenario 1: shorter tenure, less experience with LBW, and less intensive practice styles. When NICU congestion level is high, the effect estimates are consistent with the prediction in scenario 4: longer tenure, more experience with at-risk newborns, and more intensive practice styles.

5.3.2 Care Utilization and Patient Outcomes

Section 5.3.1 provide evidence that unscheduled high-risk newborn admissions affect the attending physician assignment among previously admitted newborns. Considering the large influence physicians have on medical decisions, this section reports estimated spillover effects on care utilization and patient outcomes. Table 5 summarizes coefficient estimates on length of say, total charges, and number of treatment procedures. The top panel indicates no overall difference between the treated and control group newborns. When focusing on estimates at each congestion level, the bottom panel shows that unscheduled high-risk admissions lead to a 7% increase in length of stay, a 10% increase in total charges, and an 8% (0.208/2.681) increase in number of procedures among incumbent newborns when NICU facilities are congested. To report the effects in levels, there is an increase of 0.6 (8.063×0.0731) days in length of stay, \$3653 (\$$34139 \times 0.107$) in total charges, and 0.2 in number of procedures. When NICU congestion level is low, the point estimates indicate a reduction in care utilization, although the effects are smaller and insignificant. The results on care utilization, especially the heterogeneity across congestion levels, are consistent with the findings on physician practice style. This implies that the changes in treatment intensity could be mostly driven by physician practice styles. I return to this point and discuss potential mechanisms in Section 6.

Table 5: Differences in Care Utilization during Hospital Stay

	(1)	(2)	(3)	(4)	(5)	(6)
Congestion	Length of Stay (log)		Total Charges (log)		# of Procedures	
All	0.00623		-0.00373		0.0177	
	(0.00984)		(0.0155)		(0.0472)	
Low	-0.0150	-0.0167	-0.0125	-0.0287	-0.0670	-0.0533
	(0.0250)	(0.0241)	(0.0349)	(0.0359)	(0.0573)	(0.0553)
Medium	-0.00638	-0.00510	-0.0399*	-0.0362	-0.00764	-0.0145
	(0.0138)	(0.0139)	(0.0231)	(0.0233)	(0.0752)	(0.0705)
High	0.0655***	0.0731***	0.0913**	0.107**	0.198*	0.208
	(0.0225)	(0.0264)	(0.0435)	(0.0444)	(0.109)	(0.125)
Covariate x Cgst	No	Yes	No	Yes	No	Yes
N	23791	23791	23791	23791	23791	23791
Y-mean	1.875	1.875	9.437	9.437	2.681	2.681

Standard errors in parentheses Standard errors clustered at hospital level

To further examine the spillover effects on care utilization, Table A3 reports effects on number of procedures performed on the 3rd day and later, on the 4th day and later, and on the 5th day and later.[22] Such estimates provide us with additional information of the effect timing. Focusing on the estimates in the bottom panel when NICU is congested, the increases in number of procedures persist over time and attenuate gradually, suggesting that the additional procedures performed are spread out during the hospital stay. Table A4 presents estimates on cumulative care utilization in the first year after birth.[23] Coefficient estimates show similar patterns and magnitudes compared to estimates in Table 5. This indicates that the increase in care utilization upon birth does not reduce subsequent utilization, and therefore is not a reallocation of care over time within the first year of life.

An important question to answer is whether increased care utilization leads to improvement in patient outcomes which provides a measure of productivity. Table 6 summarizes effect estimates on in-hospital mortality[24], hospital transfer, and readmission measures. None of the coefficients are significant, suggesting no clear improvement in patient outcomes. One may criticize that these outcome measures are rare events and it might be underpowered to capture any meaningful effects. However, this analysis focuses on mid-risk newborns, who have higher likelihood of having adverse health conditions than an average healthy newborn. The 1-year readmission measure reported in Table 6 has a sample average of 7%. The point estimates are insignificant and small in magnitudes. The relatively narrow confident interval indicates that we can reasonably conclude with a zero effect.

The findings of increased treatment intensity and lack of observable health benefits imply a low or zero return to the additional care utilization among incumbent newborns. It is likely that the level of care provision among mid-risk newborns has reached the "flat-of-the-curve". Therefore, the more intensive practice styles of highly specialized physicians do not generate noticeable patient benefits. One may interpret that assigning highly specialized physicians to previously admitted lower-risk newborns when hospitals are congested results

[22] Procedures performed on 3rd+ days after birth are performed after the day of unscheduled high-risk admissions for all newborns in the treated group.

[23] For patients admitted in the last year in the sample, such cumulative measure is downward biased due to sample period limitation. This issue is addressed by the time fixed effects included in the regressions.

[24] In-hospital mortality does not capture infant deaths outside hospitals. The 1-year in-hospital mortality is computed by tracing all hospital admissions of a newborn during the first year after birth. The in-hospital mortality is 0.7%, and the cumulative 1-year in-hospital mortality is 0.78% in my mid-risk sample. To benchmark the mortality measures in this paper, the overall 28-day and 1-year infant mortality in 2005-2009 U.S. Linked Birth/Infant Death Cohort Data is 0.45% and 1.4% among newborns with birth weight between 1500g to 2500g.

in wasteful medical spending. However, such a conclusion ignores the costs to achieve a seemingly more efficient physician–patient matching. When hospitals are congested, additional physicians will be needed to reduce waiting and ensure care quality. In the case of no high-risk newborn admissions, hospitals may bring in additional physicians normally treating mid-risk newborns. But with unexpected high-risk admissions, hospitals need to call in highly specialized physicians regardless. Therefore, not letting the highly specialized physicians treat previously admitted lower-risk newborns and bringing in additional physicians with better-suited practice styles will incur additional costs. Hence, when taking into account the costs in optimizing physician–patient assignment, the spillover effects and hospitals' responses may be interpreted as a constrained optimization to accommodate fluctuations in care demand.

5.3.3 Effect Heterogeneity

Summary statistics in Table 1 and Figure B7 - B10 indicate that care utilization and attending physician characteristics differ significantly across birth weight groups. If care providers consider birth weight as an important metric in making medical decisions, spillover effects of unscheduled high-risk admissions on incumbent newborns may also vary over birth weight. To flexibly trace out the distribution of spillover effects over birth weight, newborns with birth weight between 1500 grams and 3500 grams are grouped into 100-gram birth weight categories. Spillover effects at high congestion levels are estimated for each birth weight group.[25]

Figure B20 - B22 and Figure B23 - B25 present marginal treatment effects on physician profile and care utilization. Although the estimates are less precise due to small sample sizes for newborns with lower birth weight, all outcomes show similar patterns: the effects appear when birth weight drops below 2300 grams. In addition, newborns with birth weight near 1500 grams stay longer in the hospital and incur higher charges but do not experience different attending physicians practice styles, possibly because they are always treated by physicians specialized in high-risk cases. At the other end, higher birth weight groups have larger sample sizes and generate more precise estimates. Newborns at the normal birth weight range, i.e. above 2500 grams, are not affected by unscheduled high-risk admissions

[25] A regression similar to equation (3) replacing congestion indicators to birth weight group indicators is implemented in the subsample of high congestion levels. Covariate coefficients are not set to vary by birth weight groups.

Table 6: Differences in Patient Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Congestion	In-hospital Mortality		1-Year Hospital Mortality		Hospital Transfer		28-Day Readmission		1-Year Readmission	
All	0.00159		0.00110		0.000381		-0.00175		0.00703	
	(0.00239)		(0.00272)		(0.00157)		(0.00330)		(0.00647)	
Low	0.00734	0.00593	0.00633	0.00553	-0.00395	-0.00346	-0.00516	-0.00639	-0.00140	0.000309
	(0.00727)	(0.00650)	(0.00715)	(0.00648)	(0.00356)	(0.00413)	(0.00392)	(0.00495)	(0.00788)	(0.00810)
Medium	-0.00247	-0.00200	-0.00228	-0.00176	0.00125	0.00171	-0.000813	-0.000460	0.00991	0.0104
	(0.00184)	(0.00192)	(0.00191)	(0.00212)	(0.00278)	(0.00269)	(0.00565)	(0.00516)	(0.00880)	(0.00792)
High	0.00291	0.00529	0.00155	0.00314	0.00458	0.00183	0.00115	-0.000393	0.0126	0.00476
	(0.00441)	(0.00485)	(0.00434)	(0.00488)	(0.00708)	(0.00759)	(0.00640)	(0.00725)	(0.0189)	(0.0195)
Covariate x Cgst	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	23791	23791	23791	23791	23791	23791	23791	23791	23791	23791
Y-mean	0.00706	0.00706	0.00782	0.00782	0.0130	0.0130	0.0171	0.0171	0.0699	0.0699

Standard errors in parentheses

Standard errors clustered at hospital level

in all outcomes.

The distribution of spillover effects over birth weight implies that specialized physicians do not treat the average incumbent newborn. To examine how physician–patient assignment is affected among previously admitted newborns, I compare incumbent newborns treated by the specialized physicians to incumbent newborns assigned to other physicians. If we assume newborns with relatively higher risks are assigned to the specialized physicians, incumbent newborns treated by other physicians would be positively selected in their health conditions. Table A5 top panel indicates that patient observables are similar between the treated and control group newborns, the same as in Table 2. Coefficients on interaction terms in panel B separate the overall differences by whether the newborns in the treated group are attended by the physicians of unscheduled high-risk newborns or by other physicians. Treated group newborns assigned to the specialized physicians have a higher c-section rate, a lower fraction of females, and a significantly lower average birth weight. The difference in female fraction is driven by the gender difference in birth weight distribution that more male than female newborns are on the lower end of the birth weight distribution. On the other hand, treated group newborns assigned to other physicians are positively selected with a lower c-section rate and higher birth weight. Panel C reports treated–control differences focusing on a subsample of newborns in the treated group when the attending physician of subsequent unscheduled high-risk admission treats no incumbent patients. The coefficient estimates indicate that newborns in the treated group have better than average health conditions when the attending physician of subsequent unscheduled high-risk admission does not treat any incumbent patients. These results suggest that hospitals are aware of physician specialization and try to match physicians' skills to suited patient conditions to maximize productivity. In a more extreme case shown in panel C, it may be hospitals' consideration that incumbent patients are too healthy to benefit from the practice styles of physicians specialized in treating high-risk cases.

5.3.4 Robustness Checks

This section presents effect estimates under alternative specifications. I first show that estimated results are consistent and of similar magnitudes under different sets of covariates. I then test alternative control groups and explore how the results change under different definitions of treated groups. According to the baseline estimates, the spillover effects are only significant in the subsample of high NICU congestion levels. Hence, I only report robustness analyses results for the high congestion subsample.

The baseline results under regression (2) include control variables of birth delivery method, insurance type, race and gender, and birth weight. To present how patient observables affect effect estimates, Table A6 reports coefficients estimated from regression (2) adding control variables one at a time. All regressions include hospital-year, birth month, and birth day of week fixed effects. Comparing across column 1 - 5, effect estimates are robust to different patient observable controls. R^2 increases and standard errors are reduced, generating more precise point estimates. When newborn birth weight is controlled for, comparing column 4 to column 5, coefficients show smaller magnitudes but are more precisely estimated.[26]

The baseline specification categorizes newborns admitted three or more days apart from unscheduled high-risk admissions as the control group. Shown in Figure 6, newborns under this control group definition could experience hospital stays overlapped with unscheduled high-risk newborns. To reduce the influence of unscheduled high-risk admissions on control group newborns, I adopt different time cutoffs in the control group definition to test for robustness. Newborns admitted further away from unscheduled high-risk admissions (i.e. 4 or more days, 5 or more days, and 6 or more days apart from the high-risk admissions) are included in three alternative control groups. The treated groups are held the same as in the baseline specification. Table A7 summarizes effect estimates with different control groups. Column 1 reports coefficients from the baseline specification. Columns 2 - 4 report coefficients using control groups further away from the unscheduled high-risk admissions. Comparing across columns, point estimates are of comparable magnitudes. Most coefficients remain significant despite of larger standard errors due to reduced sample sizes.[26]

To explore how effect magnitudes vary with the level of hospital stay overlap with the high-risk newborns, alternative treated groups are defined to include newborn admissions with different time gaps prior to unscheduled high-risk admissions: (1) within the two days prior (baseline), (2) one day prior, (3) 0-12 hours prior,[27] (4) 12-24 hours prior, (5) 24-36 hours prior, and (6) 36-48 hours prior to unscheduled high-risk admissions. Treated group (2) is a subgroup of the baseline treated group. All newborns in the treated group

[26] Estimates on patient outcomes in Table 6 stay small and insignificant, robust to alternative specifications. Results available upon request.

[27] Newborns born within 12 hours prior to unscheduled high-risk admissions are likely to have endogeneity issues. Considering the time of labor, hospitals might be aware of the incoming high-risk newborns and pre-respond to the demand shocks.

(3) - (6) are admitted within two days prior to unscheduled high-risk admissions. However, more than 70% of newborns in the treated group (3) and around 20% of newborns in the treated group (4) are admitted on the same day of unscheduled high-risk admissions, hence they are not included in the baseline analyses. The control group is held the same as in the baseline specification. Table A8 column 1 reports the spillover effects from baseline specification. Column 2 reports spillover effects among newborns admitted one day prior to unscheduled high-risk admissions. The subgroup of mid-risk newborns admitted prior but closer to unscheduled high-risk admissions shows similar effects across all outcomes. Despite smaller treated group sample size, effects on care utilization attain similar precision levels but estimates on attending physician characteristics are less precisely estimated. Columns 3 - 6 report effect estimates on incumbent newborns by their admission time in four non-overlapping 12-hour intervals. Spillover effects on care utilization are concentrated among newborns admitted 12-24 hours prior to unscheduled high-risk admissions. Effects on attending physicians are mainly driven by newborns admitted 12-36 hours prior, but are less precisely estimated.[26]

Another possible source of variation in spillover effects is whether incumbent newborns stay inside or outside NICU. In this analysis, I focus on newborns who are directly admitted to NICU after birth and newborns who are never admitted to NICU. Since unscheduled high-risk admission might affect the NICU admission decision, newborns admitted to NICU on days after the day of birth are excluded in the analysis. The same sample definition is applied to the control group. To compare effects among NICU and non-NICU newborns, I interact an inside-NICU indicator with the treated group indicator in regression (2) and control for the baseline effect of staying inside NICU.[28] Table A9 presents estimates among the entire sample and among NICU and non-NICU patients. Coefficients in the top row differ slightly from those under the baseline specification because of the exclusion of newborns with non-immediate NICU admissions. The middle row reports effects among NICU incumbent newborns and the bottom row reports effects among non-NICU incumbent newborns. Comparing the middle and bottom rows, all spillover effects are concentrated among NICU incumbent newborns. No difference is seen in patient outcomes in either patient groups (not reported). It is worth noting that newborns with im-

[28] An alternative approach is to implement regression (2) in subsamples of NICU and non-NICU newborns. The subsample approach allows coefficients on covariates to differ inside and outside NICU. However, such approach reduces the sample size by half hence leads to imprecise estimates. Hypothetically, there is little reason to believe patient observables affect outcomes significantly different inside and outside NICU.

mediate NICU admission are in worse heath conditions compared to non-NICU newborns. Hence, it is hard to disentangle whether the heterogeneity is caused by NICU admission or newborn health conditions, as shown in Section 5.3.3.

Newborns with birth weight of 2500 grams or above have low health risks and have limited demand for care after birth. They stay mostly in regular nurseries that are physically separate from NICU facilities. Hence, unscheduled high-risk newborn admissions are expected to have little influence on low-risk newborns. Table A10 summarizes effect estimates among the low-risk sample. Only subsample estimates from equation (2) by congestion levels are reported. Comparing coefficients in Table A10 to point estimates in the bottom panel even columns of Table 3, 4, and 5, all coefficients from the low-risk sample are precisely estimated zeros. This result is consistent with the hypothesis that low-risk newborns have limited interaction with high-risk admissions even with overlapping hospital stays.

To complete the analysis of the spillover effect, I also estimate the spillover effects on mid-risk newborns admitted on the same day or within the two days after unexpected high-risk admissions in Appendix Section B.1. I also present effect estimates on incumbent newborns using c-section VLBW birth admissions as demand shocks in Appendix Section B.2.

6 Possible Mechanisms

Results in Section 5.3.2 show that newborns admitted prior to unscheduled high-risk admissions are treated by physicians with more intensive practice styles and receive increased medical care. One plausible explanation is that the increase in treatment intensity is driven by physicians' practice styles unrelated to patient conditions. Hence, little benefit is observed in newborn health.

To explore the impact of physicians' practice styles, I follow Baron and Kenny (1986) to test a mediation hypothesis. Specifically, I seek to decompose the total spillover effects into direct effects that are not associated with the attending physicians' practice styles and indirect effects that are mediated through physicians' practice styles. Partial mediation occurs if the mediator variable accounts for some, but not all, of the total spillover effects.

Denoting the measure of physicians' practice styles as variable Phy, two regression equations are estimated simultaneously following the variable notation in equation (2):

$$Phy_{i,h,t} = \phi_1 \cdot Pre_{i,h,t} + \beta X_{i,h,t} + \mu_{i,h,t},$$
$$Y_{i,h,t} = \phi_2 \cdot Pre + \theta \cdot Phy_{i,h,t} + \beta X_{i,h,t} + \epsilon_{i,h,t}.$$

ϕ_2 measures the direct effect of Pre on Y and $\phi_1 \cdot \theta$ measures the indirect effect through physicians' practice styles, i.e. the mediator variable Phy.[29] The total effect is the summation of direct and indirect effects $(\phi_1 \cdot \theta + \phi_2)$.[30]

Table A11 column 1,3, and 5 report coefficients estimated simultaneously from the above two equations using structural equation modeling (SEM). The physician practice baseline measures defined in Section 3.3 are tested as mediator variables Phy. The direct effect is measured by the regression coefficient ϕ_2. The indirect and total effects are calculated by "nonlinear combinations of estimators" function *nlcom* where the standard errors are computed using the Delta Method. Indirect % measures the indirect fraction of total effect, i.e. the fraction through mediator variable Phy. Columns 2, 4, and 6 report the total spillover effects estimated from equation (2), which differ slightly from Table 5 due to the exclusion of observations missing physician practice measures. The indirect effect percentages reported in the bottom panel imply that physicians' practice styles account for 25% - 40% of the total increases in care utilization. Considering that the mediating variable Phy only captures one dimension of physician practice, the decomposed indirect effects through physician style can be interpreted as a lower bound. There could also exist spillovers in treatment practice across physicians. These effects are closely related to physician practice, but will be captured as direct effects instead of indirect effects through physician practice style measure Phy.

To further explore how physician styles affect care utilization, I test whether procedures performed on the unscheduled high-risk newborns increase the probability of receiving the same procedures among incumbent newborns patients. Specifically, I analyze three ICD-9 procedures that high-risk newborns frequently receive on the day of admission and are also common among mid-risk newborns: 93.90 non-invasive mechanical ventilation, 99.15

[29]For patient i admitted on date t attended by physician p, $Phy_{i,h,t}$ is calculated to be average Y of patients discharged by the physician p up to date $(t-1)$.

[30]Although the Baron and Kenny (1986) method is the most commonly used approach in testing mediation effects and some associated technical issues can be addressed by structural equation modeling (SEM), it may suffer identification issues and the effect decomposition needs to be interpreted with caution (MacKinnon et al., 2007; Hayes, 2009; Zhao et al., 2010).

parenteral infusion of concentrated nutritional substances, and 99.83 other phototherapy. A regression modified from equation (2) is implemented:

$$\begin{aligned} Y^j_{i,h,t} &= \gamma_0 \cdot Pre_{i,h,t} \times (Proc^j_{i,h,t} = 0) + \gamma_1 \cdot Pre_{i,h,t} \times (Proc^j_{i,h,t} = 1) \\ &+ \beta X_{i,h,t} + \tau_{h,y} + \tau_{dow} + \tau_m + \epsilon_{i,h,t} \end{aligned} \tag{4}$$

- $Y^j_{i,h,t,s} = 1$ if patient i admitted on date t in hospital h receives procedure j during the 2 days after the admission day.
- $Pre_{i,h,t} = 1$ if patient i admitted on date t in hospital h encounters an unscheduled high-risk newborn within the 2 days the admission day.
- $Proc^j_{i,h,t} = 1$ the unscheduled high-risk newborn receives procedure j on the day of admission.
- γ_0 captures the spillover effect when newborns in the treated group encounter an unscheduled high-risk newborn who does not receiving procedure j on the day of admission.
- γ_1 captures the spillover effect when newborns in the treated group encounter an unscheduled high-risk newborn who receives procedure j on the day of admission.

Table A12 summarizes regression results on the three procedures that are common both among high-risk and mid-risk newborns. The odd columns summarize the regression results among all mid-risk newborns and the even columns report estimates at high congestion levels. Encountering an unscheduled high-risk newborn receiving one of the three procedures on the day of admission significantly increases the probability of receiving the same procedure among incumbent newborns. No similar increase in procedure use is observed when encountering an unscheduled high-risk newborn not receiving such procedure. Table A12 even columns present the estimated effects when NICUs are congested. Although the estimates are less precise due to the reduction in sample size, the point estimates are positive and take larger values compared to effect estimates when the unscheduled high-risk newborns receive no such procedures. This finding suggests a mechanism through physician practice in addition to changes in physician–patient matching: physicians' practice styles on incumbent newborns may be directly influenced by the presence of unexpected high-risk newborns. This dynamic pattern of physicians' practice styles has been shown in many medical studies that physicians may resort to the availability heuristic, i.e. previous or concurrent patient events, in making treatment decisions (Choudhry et al., 2006). One could further test for the extent of cross-physician spillovers by checking whether such mechanism mainly exists among incumbent newborns treated by the attending physician of unexpected high-risk newborns. However, such analysis is severely underpowered due to the insufficient sample size of the data.

Another possible explanation of minimal health improvement from the increased treatment intensity is that unscheduled high-risk admissions may have led to initial negative spillovers on previously admitted newborn patients. Hence, hospitals retain incumbent newborns for longer stays and/or treat them more intensively to compensate for the initial negative effects. To test the crowding-out hypothesis, I investigate whether unscheduled high-risk admissions lead to delays in standard procedures after birth such as vaccinations and hearing tests, the two most common procedures after birth. Table A13 report regression estimates from equation (2) in the entire mid-risk sample and by congestion levels. Columns 1 and 2 report linear probability model estimates. The top panel sample averages show that more than 60% and 40% of newborns receive hearing tests and vaccination before discharge. To investigate any possible delays, I adopt two measures in columns 3 - 6, the number of days before receiving the two procedures, and a binary variable concerning whether the procedure is performed within the first 4 days during the stay. If initial crowding-out occurs, we would expect a longer period before receiving the procedure or a lower probability of receiving the procedure within the first few days. The sample averages in columns 5 - 6 indicate that hearing tests and vaccination are performed typically on the 3rd and 4th day after birth. Since newborns in the treated group are admitted within the two days prior to unscheduled high-risk admissions, these two procedures are subject to any crowding-out effects if that exists. All regression coefficients in Table A13 are small and insignificant, indicating that unscheduled high-risk admissions do not change the overall probability or result in initial delays of receiving these two common procedures among incumbent newborns. Hence, it is unlikely that initial crowding-out masks any health benefits from increased care utilization among newborns admitted prior to unscheduled high-risk admissions.

7 Conclusion

The increase in labor specialization and the expansion of professional service-oriented sectors have increased the importance of skill–task matching in advanced economies. Typically, a specialist outperforms a jack-of-all-trades in tasks that require specific knowledge and extensive training. However, it is has been established theoretically that the need to coordinate specialized activities might limit and even outweigh the benefits in production brought by specialization (Becker and Murphy, 1992; Garicano, 2000; Dessein and Santos, 2006; Fuchs and Garicano, 2010). This paper is the first to provide empirical evidence on firms' skill–task matching behaviors in response to demand uncertainty. When demand is unpredictable, matching experts' skills to tasks they are suited for requires frequent and recurring decision-making – an often-overlooked challenge in modern production. Understanding firms' decisions under such condition is essential because it will provide researchers and policy-makers with key information in efficiency evaluation.

The health care sector provides an appealing research setting for examining skill–task matching under demand uncertainty. Hospitals cannot fully control the arrivals of patients, and the assignments of patients to physicians are clearly defined in hospital discharge records. This paper capitalizes on the availability of this rich information, using discharge data from New York City hospitals to study the skill–task matching behaviors in response to demand shocks arising from unscheduled high-risk admissions. Additionally, decisions in the health care sector are expensive: not only do they drive high aggregate medical spending, but they also affect patient morbidity and mortality. The high costs involved make the subject of this study a practically meaningful area for research inquiry and policy-making.

Empirical findings in this paper show that hospitals summon highly specialized physicians and reoptimize physician–patient assignment upon temporary increases in care demand. This leads to spillover effects on patients admitted prior to unscheduled high-risk admissions: when hospitals are congested, these incumbent patients are more likely to be attended by physicians with more intensive practice styles who specialize in treating high-risk cases, leading to increases in care utilization for these patients without any detectable improvement in outcomes. The low productivity of specialized physicians when performing less familiar tasks has important implications. Whereas it seems almost certain that less specialized individuals would not perform as well as highly specialized experts at complex tasks, more specialized or highly trained experts are *not* better at all tasks. Instead, experts' productivity strongly depends on their task assignments, making good skill–task matching essential in highly specialized production.

Hospitals' responses to the unexpected demand provide broader insights into industries with a high level of specialization. The finding that hospitals are more likely to allow skill–task mismatch when matching is more costly suggests that the costs of assigning tasks are carefully considered in health care production. Meanwhile, many specialized industries face similar challenges where evaluating the cost-benefit trade-offs in managing specialized labor inputs is not straightforward, especially under demand fluctuations. Additional empirical studies focusing on demand uncertainty in analyzing firms' task assignment behaviors would add great value to the literature and provide useful evidence for policy-makers as they design job assignment schemes. Although the research setting in this paper is not suited for determining the optimal level of skill–task mismatch in production, the findings underscore the need to carefully consider all costs associated with task assignment when evaluating efficiency. These findings provide vital information for policy-makers looking to identify waste in utilization and create incentives to enhance efficiency.

References

AAMC. Physician specialty data report - Association of American Medical Colleges (AAMC), 2018. URL https://www.aamc.org/download/493090/data/2018-aamc-physician-specialty-data-report.pdf. Acccessed: 23-May-2019.

Douglas Almond and Joseph J Doyle. After midnight: A regression discontinuity design in length of postpartum hospital stays. *American Economic Journal: Economic Policy*, 3(3):1–34, 2011.

Douglas Almond, Joseph J Doyle Jr, Amanda E Kowalski, and Heidi Williams. Estimating marginal returns to medical care: Evidence from at-risk newborns. *The quarterly journal of economics*, 125(2):591–634, 2010.

Katherine Baicker and Amitabh Chandra. Medicare spending, the physician workforce, and beneficiaries' quality of care: Areas with a high concentration of specialists also show higher spending and less use of high-quality, effective care. *Health Affairs*, 23 (Suppl1):W4–184, 2004.

Laurence C Baker, Ciaran S Phibbs, Cassandra Guarino, Dylan Supina, and James L Reynolds. Within-year variation in hospital utilization and its implications for hospital costs. *Journal of health economics*, 23(1):191–211, 2004.

Christopher Barbey, Nikhil Sahni, Robert Kocher, and Michael Chernew. Physician workforce trends and their implications for spending growth. *Health Affairs Blog*, 2017. doi: 10.1377/hblog20170728.061252.

Reuben M Baron and David A Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173, 1986.

Ann P Bartel, Ciaran S Phibbs, Nancy Beaulieu, and Patricia Stone. Human capital and organizational performance: Evidence from the healthcare sector. Technical report, National Bureau of Economic Research, 2011.

Gary S Becker and Kevin M Murphy. The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, 107(4):1137–1160, 1992.

Prashant Bharadwaj, Katrine Vellesen Løken, and Christopher Neilson. Early life health interventions and academic achievement. *American Economic Review*, 103(5):1862–91, 2013.

Amitabh Chandra and Douglas O Staiger. Productivity spillovers in health care: evidence from the treatment of heart attacks. *Journal of Political Economy*, 115(1):103–140, 2007.

Amitabh Chandra and Douglas O Staiger. Identifying sources of inefficiency in health care. Technical report, National Bureau of Economic Research, 2017.

Amitabh Chandra, David Cutler, and Zirui Song. Who ordered that? the economics of treatment choices in medical care. In *Handbook of health economics*, volume 2, pages 397–432. Elsevier, 2011.

Niteesh K Choudhry, Geoffrey M Anderson, Andreas Laupacis, Dennis Ross-Degnan, Sharon-Lise T Normand, and Stephen B Soumerai. Impact of adverse events on prescribing warfarin in patients with atrial fibrillation: matched pair analysis. *Bmj*, 332 (7534):141–145, 2006.

Eric Chyn, Samantha Gold, and Justine S Hastings. The returns to early-life interventions for very low birth weight children. Technical report, National Bureau of Economic Research, 2019.

Janet Currie. Healthy, wealthy, and wise: Is there a causal relationship between child health and human capital development? *Journal of Economic Literature*, 47(1):87–122, 2009.

Janet Currie and W Bentley MacLeod. Diagnosing expertise: Human capital, decision making, and performance among physicians. *Journal of labor economics*, 35(1):1–43, 2017.

Janet Currie, W Bentley MacLeod, and Jessica Van Parys. Provider practice style and patient health outcomes: the case of heart attacks. *Journal of health economics*, 47: 64–80, 2016.

David Cutler, Jonathan Skinner, Ariel Dora Stern, and David Wennberg. Physician beliefs and patient preferences: a new look at regional variation in health care spending. Technical report, National Bureau of Economic Research, 2013.

Min Dai, Jussi Keppo, and Tim Maull. Hiring, firing, and relocation under employment protection. *Journal of Economic Dynamics and Control*, 56:55–81, 2015.

Wouter Dessein and Tano Santos. Adaptive organizations. *Journal of Political Economy*, 114(5):956–995, 2006.

Joseph Doyle. Measuring physician quality: Evidence from physician availability. *Working Paper*, 2018.

Joseph J Doyle. Returns to local-area health care spending: evidence from health shocks to patients far from home. *American Economic Journal: Applied Economics*, 3(3):221–43, 2011.

Joseph J Doyle, Steven M Ewer, and Todd H Wagner. Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of health economics*, 29 (6):866–882, 2010.

Joseph J Doyle, John A Graves, Jonathan Gruber, and Samuel A Kleiner. Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, 123(1):170–214, 2015.

Joseph J Doyle, John A Graves, and Jonathan Gruber. Uncovering waste in us healthcare: Evidence from ambulance referral patterns. *Journal of health economics*, 54:25–39, 2017.

Todd Elder, David N Figlio, Scott A Imberman, and Claudia Persico. The role of neonatal health in the incidence of childhood disability. Technical report, National Bureau of Economic Research, 2019.

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

Itzik Fadlon and Jessica N Van Parys. Primary care physician practice styles and patient care: Evidence from physician exits in medicare. Technical report, National Bureau of Economic Research, 2019.

David Figlio, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth. The effects of poor neonatal health on children's cognitive development. *The American Economic Review*, 104(12):3921–3955, 2014.

Elliott S Fisher, David E Wennberg, Thrse A Stukel, Daniel J Gottlieb, F Lee Lucas, and Etoile L Pinder. The implications of regional variations in medicare spending. part 1: the content, quality, and accessibility of care. *Annals of internal medicine*, 138(4):273–287, 2003a.

Elliott S Fisher, David E Wennberg, Thrse A Stukel, Daniel J Gottlieb, F Lee Lucas, and Etoile L Pinder. The implications of regional variations in medicare spending. part 2: health outcomes and satisfaction with care. *Annals of internal medicine*, 138(4):288–298, 2003b.

Seth Freedman. Capacity and utilization in health care: the effect of empty beds on neonatal intensive care admission. *American Economic Journal: Economic Policy*, 8(2):154–85, 2016.

William Fuchs and Luis Garicano. Matching problems with expertise in firms and markets. *Journal of the European Economic Association*, 8(2-3):354–364, 2010.

Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of political economy*, 108(5):874–904, 2000.

Robert Gibbons and Michael Waldman. Careers in organizations: Theory and evidence. *Handbook of labor economics*, 3:2373–2437, 1999.

Andrew F Hayes. Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Communication monographs*, 76(4):408–420, 2009.

HCUP. Most common diagnoses for inpatient stays - Healthcare Cost and Utilization Project (HCUP), 2015. URL https://www.hcup-us.ahrq.gov/faststats/NationalDiagnosesServlet. Acccessed: 12-April-2019.

Thomas Hoe. Does hospital crowding matter? evidence from trauma and orthopedics in england. 2018.

Nathan R Hoot and Dominik Aronsky. Systematic review of emergency department crowding: causes, effects, and solutions. *Annals of emergency medicine*, 52(2):126–136, 2008.

David Hughes and Alistair McGuire. Stochastic demand, production responses and hospital costs. *Journal of health economics*, 22(6):999–1010, 2003.

David P MacKinnon, Amanda J Fairchild, and Matthew S Fritz. Mediation analysis. *Annu. Rev. Psychol.*, 58:593–614, 2007.

Mindy Marks and Moonkyung Kate Choi. Baby boomlets and baby health: Hospital crowdedness, hospital spending, and infant health. *American Journal of Health Economics*, pages 1–48, 2018.

David Molitor. The evolution of physician practice styles: evidence from cardiologist migration. *American Economic Journal: Economic Policy*, 10(1):326–56, 2018.

OECD. Health spending, 2019. URL https://data.oecd.org/healthres/health-spending.htm. Acccessed: 23-Sep-2019.

Charles E Phelps. *Health economics*. Routledge, 2016.

Rajiv Sharma, Miron Stano, and Renu Gehring. Short-term fluctuations in hospital demand: implications for admission, discharge, and discriminatory behavior. *The Rand journal of economics*, 39(2):586–606, 2008.

Yu-Chu Shen and Renee Y Hsia. Ambulance diversion associated with reduced access to cardiac technology and increased one-year mortality. *Health Affairs*, 34(8):1273–1280, 2015.

Jonathan Skinner and Elliott Fisher. Reflections on geographic variations in us health care. *Dartmouth Institute for Health Policy and Clinical Practice, May*, 12, 2010.

Celeste Torio and Brian Moore. National inpatient hospital costs: the most expensive conditions by payer, 2013: statistical brief# 204. 2016.

Xinshu Zhao, John G Lynch Jr, and Qimei Chen. Reconsidering baron and kenny: Myths and truths about mediation analysis. *Journal of consumer research*, 37(2):197–206, 2010.

A Stylized Model

Consider a simple production model with discrete task types $i = 1, ..., n$ and expert types $j = 1, ..., n$. To model hospitals' responses upon patient admission, I assume that the content and arrival time of tasks (patients) are exogenous and the available experts (physicians) on site are given. In a long-run general equilibrium model, hospitals may be able to adjust both the demand and supply by selectively admitting patients or employing physicians of certain types. However, in the short-run partial equilibrium model I explore here, the assumptions of exogenous patient demand and fixed physician supply do not seem overly restrictive.

In the production function, I follow Chandra and Staiger (2017) and focus on patient health benefits which include any reduction in mortality or morbidity from medical care. Although hospital decisions could be influenced by other incentives such as financial returns, many studies show that physicians care about patient outcomes. In addition, if we assume physician experts only differ by their comparative advantage in skills, i.e. horizontal differentiation, there is insufficient reason to assume *ex-ante* that some physicians will generate higher financial revenue or costs than others at the task assignment stage. For a similar reason, I ignore any actual costs of treatment and only focus on the labor cost in the cost function. Formally, we assume that

$$Y_i = \sum_{j=1}^{n} \lambda_{ij} L_j,\ C_i = \sum_{j=1}^{n} \omega_{ij} L_j,$$

where Y_i and C_i are the health outcome and labor cost associated with a patient of type i. L_j is the input unit of type j physician assigned to the patient. Under a one-to-one matching, only one of L_j takes a unit value and the rest are zero,

Under horizontal differentiation, the model assumes homogeneous marginal product λ and marginal cost ω of physicians when tasks are randomly assigned. Marginal product increases if a good skill–task matching is achieved, i.e. $i = j$. This in turn incurs extra matching costs.

$$\lambda_{ij} = \begin{cases} \lambda_H, & i = j \\ \lambda_L, & i \neq j \end{cases}, \text{ and } \omega_{ij} = \begin{cases} c + M, & i = j \\ c, & i \neq j \end{cases}.$$

Upon the admission of a type i patient, hospitals assign one unit of physician input and maximize the net benefit $(Y_i - C_i)$ by choosing among physician types:

$$Max_j \sum_{j=1}^{n} (\lambda_{ij} - \omega_{ij}) L_j.$$

Hospitals prioritize matching, i.e. $j^* = i$, if $(\lambda_H) - c - M > \lambda_L - c$. The optimal decision depends on whether the cost of matching M outweighs its return $(\lambda_H - \lambda_L)$.

The matching cost M, although enters the model as a constant, can vary significantly depending on environmental factors. When hospitals are congested, there may be little flexibility in resource allocation. Hence the matching cost will be high and it will be inefficient for hospitals to prioritize matching. If we assume the marginal cost c takes a U-shape over physicians' existing workload, i.e. high costs to call in new physicians or assign patients to fully-occupied physicians, achieving a good matching will incur a high opportunity cost when some "$j \neq i$" type physicians are on site with excess capacity but no "$j = i$" type physician is readily available. The level of differences between patients affects the magnitude of $(\lambda_H - \lambda_L)$. If a type i patient does not differ much from other patient types, the cost of mismatch will be low and hospitals will be reluctant to prioritize physician–patient matching. Moreover, uncertainty in demand will lead to additional inertia in adjusting labor input that the optimal strategy allows certain level of mismatch due to the high costs in constantly tracing the optimal assignment. (Dai et al., 2015).[31]

B Spillover Effects - Supplementary Results

B.1 Newborns Admitted on the Same Day or After Unscheduled High-risk Admissions

Newborns admitted on the same day or soon after unscheduled high-risk admissions also experience considerable overlaps in hospital stay with the high-risk newborns. I repeat the analyses following regression equation (2) to estimate spillover effects at each con-

[31] The stochastic control model in Dai et al. (2015) assumes that the demand follows a Brownian motion and the labor control solution is solved based on Hamilton-Jacobi-Bellman equation (similar models are widely adopted in finance for option pricing). Their model predicts a "no-action region" where a firm does not constantly trace the optimal labor input level and only adjusts labor input when the labor-to-demand ratio hits an upper or lower bound (see Figure B26).

gestion level on mid-risk newborns admitted on the same day or in the two days post to unscheduled high-risk admissions. Section 5.2 discusses the potential selection issue that hospitals may be selective in admitting newborns in the presence of unscheduled high-risk admissions. Table A14 show marginal differences in observables between newborns admitted in the two days post to unscheduled high-risk admissions and the control group. The differences between newborns admitted on the same day and the control group are more salient. Even in absent of observable differences, it is acknowledged in the literature that selection in patient unobservables is plausible. Given that high-risk admissions consume a large amount of hospital resources, it is likely that hospitals respond in subsequent admission decisions. Hence, any estimated spillover effects among newborns admitted on the same day or after unscheduled high-risk admissions need to be interpreted with caution.

Table A16 summarizes effect estimates among newborns admitted during the two days post to unscheduled high-risk admissions. Opposite to the findings on incumbent newborns, newborns admitted soon after are assigned to physicians with less intensive treatment styles when NICU facilities are congested. However, the difference in physician style does not lead to reduction in treatment intensity. Little difference is seen in care utilization or patient outcomes reported in the bottom panel. With the concern that hospitals might selectively admit healthier newborns after unscheduled high-risk admissions, it's less clear whether the differences in physicians' practice styles could be interpreted as spillover effect or a result of selection in patient admissions.

Table A17 summarizes effect estimates among newborns admitted on the same day of any unscheduled high-risk admissions. The top panel suggests that newborns are treated by physicians more specialized in treating high-risk births with more intensive styles when NICU facilities are not congested. The bottom panel indicates that care utilization is marginally higher when NICU occupancy is either low or high. In addition, the in-hospital mortality is lower when NICU congestion is at medium level. In general, the findings on care utilization and patient outcomes do not point to any internally consistent patterns. Splitting newborns admitted on the same day by whether admitted prior to or after the unscheduled high-risk admissions does not eliminate covariates imbalance or produce more consistent effect estimates. Since the mother of the high-risk newborn could be in labor hours before the admission time of the baby, it is hard to determine the time horizon on the day of unscheduled admissions during which hospital admission decisions are less likely to be affected by the high-risk newborn.

B.2 Newborns Admitted Prior to C-section High-risk Admissions

C-section high-risk births also lead to increases in care demand in hospitals. They can be either scheduled or emergent (unscheduled). However, such information is not available in the data. Figure 1 - B1 and the LASSO prediction model estimates in Table A19 suggest that some c-section high-risk admissions may be scheduled, as discussed in Section 3.1. If hospitals and physicians are expecting the increase in care demand resulted from c-section high-risk newborn admissions, the spillover effects would be attenuated or nonexistent. It is also possible that obstetricians intentionally schedule other prior c-section births to accommodate the c-section high-risk births which will violate the exogeneity assumption.

Table A15 presents covariates differences between newborns admitted in the two days prior to c-section high-risk admissions and newborns admitted three or more days away by congestion levels. The treated and control group newborns are similar in observables, except that the treated group has a slightly lower female fraction and a slightly higher Medicaid fraction at low and medium congestion levels. Table A18 summarizes spillover effects of c-section high-risk admissions on incumbent newborns at each congestion level using regression (2). unlike in the cases of unscheduled high-risk admissions, there is little effect on physician characteristics when NICU facilities are congested. If the previous conjecture is correct that c-section high-risk admissions are expected, hospitals should arrange staff so that assignment of attending physicians among other patients are not affected. Consistent with estimates on attending physician characteristics, no increase in care utilization is seen when NICU facilities are congested. Compared to estimates in Table 5, point estimates in Table A18 are more precise but much smaller in magnitude. These is a reduction in out-transfer rate when the congestion level is high. It could be that hospitals are better staffed which in turn benefits incumbent newborns and reduces the need of out-transfer.

C Appendix Tables

Table A1: Sample Partition

Birth Weight Categories	< 1500g High-risk	$[1500, 2500)$g Mid-risk	≥ 2500g Low-risk
Vaginal Births	Demand Shocks (main analyses)	Analysis Sample	Placebo Sample
C-section Births	Demand Shocks (supplementary analyses)		

Table A2: Differences in Attending Physician Practice Style[a]

	(1)	(2)	(3)	(4)	(5)	(6)
Congestion	Avg. Res. Len. of Stay (log)		Avg. Res. Charges (log)		Avg. Res. # of Proc.	
All	0.00221		0.00555		0.0181*	
	(0.00355)		(0.00993)		(0.0101)	
Low	-0.00654	-0.00942	-0.0156	-0.0223	-0.00527	-0.0125
	(0.00611)	(0.00643)	(0.0166)	(0.0161)	(0.0239)	(0.0262)
Medium	0.00492	0.00550	0.0139	0.0142	0.0255	0.0267
	(0.00378)	(0.00388)	(0.0113)	(0.0105)	(0.0194)	(0.0200)
High	0.00809	0.00784	0.0159	0.0183	0.0323	0.0309
	(0.00630)	(0.00619)	(0.0170)	(0.0159)	(0.0204)	(0.0197)
Covariate x Cgst	No	Yes	No	Yes	No	Yes
N[b]	23535	23535	23535	23535	23535	23535
Y-mean	0.0622	0.0622	0.201	0.201	0.213	0.213

Standard errors in parentheses　　Standard errors clustered at hospital level

[a] For patients admitted on day t attended by physician p, physician practice style measures are defined to be residual averages among newborn patients discharged up to day t-1 by physician p. Residuals are generated by controlling for birth hospital-year, birth month, birth day of week, birth delivery method, insurance type, race, gender, and birth weight.
[b] A small fraction of patients have missing physician practice measures because there is no previously discharged patients by their attending physicians.

Table A3: Differences in Number of Procedures

	(1)	(2)	(3)	(4)	(5)	(6)
Congestion	# of Proc (3rd+ day)		# of Proc (4th+ day)		# of Proc (5th+ day)	
All	0.0134		0.0103		0.0205	
	(0.0365)		(0.0236)		(0.0266)	
Low	-0.0544	-0.0373	-0.0370	-0.0190	-0.0153	0.0146
	(0.0637)	(0.0528)	(0.0532)	(0.0411)	(0.0397)	(0.0308)
Medium	-0.00591	-0.00736	-0.0148	-0.0188	-0.00562	-0.00792
	(0.0486)	(0.0513)	(0.0329)	(0.0347)	(0.0365)	(0.0369)
High	0.161*	0.160*	0.141**	0.144**	0.136*	0.125*
	(0.0850)	(0.0832)	(0.0694)	(0.0682)	(0.0749)	(0.0710)
Covariate x Cgst	No	Yes	No	Yes	No	Yes
N	23791	23791	23791	23791	23791	23791
Y-mean	1.102	1.102	0.769	0.769	0.597	0.597

Standard errors in parentheses Standard errors clustered at hospital level

Table A4: Differences in One-Year Care Utilization

	(1)	(2)	(3)	(4)	(5)	(6)
Congestion	1-Year Total Len. of Stay (log)		1-Year Total Charges (log)		1-Year Total # of Proc.	
All	0.00707		-0.00251		0.0304	
	(0.00999)		(0.0168)		(0.0534)	
Low	-0.0150	-0.0191	-0.00868	-0.0287	-0.0680	-0.0328
	(0.0201)	(0.0209)	(0.0287)	(0.0317)	(0.0601)	(0.0633)
Medium	-0.00465	-0.00160	-0.0422*	-0.0354	0.00314	-0.000685
	(0.0149)	(0.0145)	(0.0247)	(0.0234)	(0.0716)	(0.0659)
High	0.0653**	0.0672**	0.0975*	0.104*	0.236	0.227
	(0.0270)	(0.0298)	(0.0514)	(0.0536)	(0.144)	(0.177)
Covariate x Cgst	No	Yes	No	Yes	No	Yes
N	23791	23791	23791	23791	23791	23791
Y-mean	1.915	1.915	9.502	9.502	2.837	2.837

Standard errors in parentheses Standard errors clustered at hospital level

Table A5: Treated - Control Differences in Newborn Characteristics

	(1) C-section	(2) Birth Weight	(3) Female	(4) White	(5) Black	(6) Medicaid
Panel A: Overall Treated - Control Differences (same as in Table 2)						
Pre	-0.0152	-5.043	-0.00888	-0.00481	-0.00363	0.00387
	(0.0123)	(6.373)	(0.00822)	(0.00957)	(0.0115)	(0.00760)
N	23791	23791	23791	23791	23791	23791
Panel B: Treated Group Newborns Assigned to the Specialized Physician vs. Other Physicians						
$Pre \times (Spec = 1)$	0.0385**	-81.07***	-0.0569***	-0.0108	0.0250	0.0224
	(0.0165)	(16.05)	(0.0150)	(0.0119)	(0.0178)	(0.0269)
$Pre \times (Spec = 0)$	-0.0372**	26.05**	0.0108	-0.00236	-0.0153	-0.00372
	(0.0170)	(10.66)	(0.0117)	(0.0129)	(0.0152)	(0.0137)
N	23791	23791	23791	23791	23791	23791
Panel C: Treated Group Newborns When Specialized Physicians Treat No Prior Patients						
Pre	-0.0351*	17.77*	0.0152	-0.00882	-0.0137	-0.00959
	(0.0173)	(9.931)	(0.0125)	(0.0119)	(0.0160)	(0.0131)
N	23008	23008	23008	23008	23008	23008

Standard errors in parentheses   Standard errors clustered at hospital level

[a] Hospital-year, month, and day of week fixed effects are included in measuring treated–control differences.
[b] Panel A reports overall differences between the treated and control group newborns. Panel B separately reports treated–control differences for incumbent newborns who are assigned to the specialized physicians and incumbent newborns who are assigned to other physicians. Panel C reports treated–control differences for a subsample of newborns in the treated group when the attending physician of subsequent high-risk admission treats no incumbent patients.

Table A6: Alternative Specifications
(Congestion = High)

	(1)	(2)	(3)	(4)	(5)
Length of Stay (log)	0.0906**	0.102**	0.103**	0.110***	0.0731***
	(0.0405)	(0.0393)	(0.0391)	(0.0382)	(0.0264)
R^2	[0.041]	[0.102]	[0.111]	[0.123]	[0.449]
Total Charges (log)	0.129	0.147*	0.148*	0.163**	0.107**
	(0.0782)	(0.0726)	(0.0732)	(0.0717)	(0.0444)
R^2	[0.121]	[0.173]	[0.179]	[0.193]	[0.484]
# of Procedures	0.237*	0.256*	0.257*	0.286**	0.208
	(0.139)	(0.134)	(0.135)	(0.140)	(0.125)
R^2	[0.179]	[0.200]	[0.204]	[0.236]	[0.359]
Experience with VLBW	0.00631*	0.00668*	0.00655*	0.00709**	0.00606*
	(0.00339)	(0.00335)	(0.00337)	(0.00336)	(0.00314)
R^2	[0.171]	[0.177]	[0.181]	[0.184]	[0.248]
Physician Avg Len. of Stay	0.0437*	0.0467**	0.0456*	0.0497**	0.0409**
(log)	(0.0224)	(0.0221)	(0.0228)	(0.0220)	(0.0188)
R^2	[0.156]	[0.164]	[0.170]	[0.177]	[0.274]
Physician Avg Total Charges	0.0767*	0.0830*	0.0805*	0.0886**	0.0708**
(log)	(0.0430)	(0.0421)	(0.0430)	(0.0414)	(0.0346)
R^2	[0.226]	[0.234]	[0.242]	[0.248]	[0.342]
Physician Avg # of Procedure	0.0896**	0.0952**	0.0937**	0.0990**	0.0849**
	(0.0382)	(0.0380)	(0.0386)	(0.0391)	(0.0348)
R^2	[0.436]	[0.441]	[0.443]	[0.446]	[0.498]
Delivery Method	No	Yes	Yes	Yes	Yes
Insurance Type	No	No	Yes	Yes	Yes
Race & Gender	No	No	No	Yes	Yes
Birth Weight (250g bin)	No	No	No	No	Yes
N	4986	4986	4986	4986	4986

Standard errors in parentheses, R^2 in square brackets Standard errors clustered at hospital level

[a] All regressions include hospital-year, birth month, and day of week fixed effects

Table A7: Alternative Control Groups
(Congestion = High)

	(1)	(2)	(3)	(4)
Length of Stay (log)	0.0731***	0.0738***	0.0788***	0.0784***
	(0.0264)	(0.0247)	(0.0249)	(0.0260)
Total Charges (log)	0.107**	0.111**	0.128**	0.132**
	(0.0444)	(0.0469)	(0.0520)	(0.0548)
# of Procedures	0.208	0.171	0.177	0.212*
	(0.125)	(0.116)	(0.120)	(0.122)
Experience with VLBW	0.00606*	0.00571	0.00696*	0.00728*
	(0.00314)	(0.00362)	(0.00366)	(0.00368)
Physician Avg Len. of Stay (log)	0.0409**	0.0369	0.0469*	0.0483*
	(0.0188)	(0.0226)	(0.0245)	(0.0255)
Physician Avg Total Charges (log)	0.0708**	0.0661	0.0850*	0.0886*
	(0.0346)	(0.0421)	(0.0451)	(0.0456)
Physician Avg # of Procedures	0.0849**	0.0772*	0.0930**	0.100**
	(0.0348)	(0.0394)	(0.0427)	(0.0449)
Control Group[a]	3+ Days Apart	4+ Days Apart	5+ Days Apart	6+ Days Apart
N	4986	4598	4250	3978

Standard errors in parentheses Standard errors clustered at hospital level

[a] Alternative control groups are defined in Section 5.3.4.

Table A8: Alternative Treated Groups
(Congestion = High)

	(1)	(2)	(3)	(4)	(5)	(6)
Length of Stay (log)	0.0731***	0.0773***	0.0662	0.168***	-0.0266	0.0240
	(0.0264)	(0.0269)	(0.0755)	(0.0549)	(0.0615)	(0.0491)
Total Charges (log)	0.107**	0.110**	0.121	0.243***	0.00886	0.00991
	(0.0444)	(0.0488)	(0.123)	(0.0819)	(0.133)	(0.0675)
# of Procedures	0.208	0.227	0.424	0.657***	-0.145	0.0947
	(0.125)	(0.140)	(0.276)	(0.202)	(0.151)	(0.192)
Experience with VLBW	0.00606*	0.00575	-0.00119	0.00663*	0.0113	-0.00453
	(0.00314)	(0.00407)	(0.00536)	(0.00388)	(0.00899)	(0.00491)
Physician Avg Len. of Stay	0.0409**	0.0530	-0.0145	0.0384	0.0920	-0.0516
(log)	(0.0188)	(0.0337)	(0.0587)	(0.0398)	(0.0671)	(0.0339)
Physician Avg Total Charges	0.0708**	0.0956	-0.0182	0.0470	0.190	-0.0910
(log)	(0.0346)	(0.0611)	(0.118)	(0.0739)	(0.134)	(0.0736)
Physician Avg # of Procedures	0.0849**	0.103	-0.000584	0.0726	0.151	-0.114*
	(0.0348)	(0.0644)	(0.102)	(0.0694)	(0.107)	(0.0560)
Treated Group[a]	1-2 Days	1 Day	0-12 Hrs	12-24 Hrs	24-36 Hrs	36-48 Hrs
N	4986	4768	4670	4661	4679	4657

Standard errors in parentheses — Standard errors clustered at hospital level

[a] Alternative treated groups are defined in Section 5.3.4.

Table A9: Effects among Patients inside and outside NICU
(Congestion = High)

	(1) Len. of Stay (log)	(2) Total Charges (log)	(3) # of Procedures	(4) Experience with VLBW	(5) Avg. Len. of Stay (log)	(6) Avg. Total Charges (log)	(7) Avg. # of Procedures
All	0.0695**	0.118**	0.200	0.00657*	0.0522**	0.0964**	0.0997**
	(0.0293)	(0.0490)	(0.127)	(0.00376)	(0.0216)	(0.0400)	(0.0386)
Inside NICU	0.104**	0.181***	0.302*	0.0100	0.0667	0.112	0.137
	(0.0471)	(0.0500)	(0.165)	(0.00842)	(0.0445)	(0.0873)	(0.0866)
Outside NICU	-0.0165	-0.0770*	-0.0288	-0.00127	-0.00345	-0.00470	-0.00532
	(0.0260)	(0.0408)	(0.0943)	(0.00520)	(0.0367)	(0.0739)	(0.0662)

Standard errors in parentheses — Standard errors clustered at hospital level

[a] Inside NICU Sample: Newborns who are directly admitted to NICU after birth.
[a] Outside NICU Sample: Newborns who are never admitted to NICU during the hospital stay.

Table A10: Placebo Test - Low-risk Newborn Sample

Congestion	(1) Len. of Stay (log)	(2) Total Charges (log)	(3) # of Procedures	(4) Physician Tenure	(5) Experience with VLBW	(6) Avg. Total Charges (log)
Low	0.00334	0.00613	0.0117	-0.0952	0.000561	0.0122
	(0.00227)	(0.00600)	(0.0110)	(0.131)	(0.000487)	(0.0120)
Medium	-0.00177	-0.00281	-0.000322	0.117	-0.000354	-0.00231
	(0.00229)	(0.00510)	(0.00794)	(0.213)	(0.000293)	(0.00999)
High	0.00165	-0.00342	-0.000951	0.212	-0.000720	-0.00959
	(0.00403)	(0.00875)	(0.0113)	(0.254)	(0.000572)	(0.0127)
N	398562	398562	398562	389269	396117	394191
Y-mean	1.257	8.269	1.524	16.77	0.0126	8.666

Standard errors in parentheses. Standard errors clustered at hospital level

[a] Sample: Singleton newborns with birth weight of 2500 grams or above.

Table A11: Mediation Analysis - Physician Practice Style (Congestion = High)

	(1)	(2)	(3)	(4)	(5)	(6)
	Length of Stay (log)		Total Charges (log)		# of Procedures	
$Phy = \alpha_1 + \phi_1 Pre + +\beta X + \mu$						
Pre	0.0409**		0.0708**		0.0849**	
	(0.0187)		(0.0345)		(0.0347)	
$Y = \alpha_2 + \phi_2 Pre + \theta Phy + \beta X + \epsilon$						
Pre	0.0590**	0.0780**	0.0686*	0.111**	0.157	0.217**
	(0.0261)	(0.0308)	(0.0408)	(0.0544)	(0.122)	(0.0995)
Phy	0.464***		0.593***		0.714***	
	(0.0311)		(0.0380)		(0.0478)	
Effect Decomposition						
Direct Effect	0.0590**		0.0686*		0.157	
(ϕ_2)	(0.0261)		(0.0408)		(0.122)	
Indirect Effect	0.0190**		0.0420**		0.0607**	
$(\phi_1 \cdot \theta)$	(0.00882)		(0.0207)		(0.0260)	
Total Effect	0.0780***		0.111**		0.217*	
$(\phi_2 + \phi_1 \cdot \theta)$	(0.0262)		(0.0432)		(0.126)	
Indirect %	0.243		0.380		0.279	
N	4942	4942	4942	4942	4942	4942

Standard errors in parentheses. Standard errors clustered at hospital level

[a] For each outcome Y, the physician practice measure Phy is the physician baseline measure of average Y defined in Section 3.3, i.e. the average Y of patients discharged by the physician p up to date $(t-1)$.
[b] The standard errors of the indirect and total effects are calculated using the Delta method.

Table A12: Treatment Procedure Spillover

	(1)	(2)	(3)	(4)	(5)	(6)
Procedure	Ventilation		Nutrition		Phototherapy	
$Pre \times (Proc^j = 0)$	-0.00243	-0.00658	-0.00993	-0.00876	-0.0104	0.00196
	(0.00390)	(0.00807)	(0.00691)	(0.0145)	(0.00768)	(0.0160)
$Pre \times (Proc^j = 1)$	0.0253**	0.0103	0.0961***	0.103*	0.147***	0.115
	(0.0109)	(0.0244)	(0.0312)	(0.0529)	(0.0324)	(0.0775)
N	23791	4986	23791	4986	23791	4986
Y-mean	0.0217	0.0231	0.0741	0.0642	0.133	0.125
Congestion	All	High	All	High	All	High

Standard errors in parentheses Standard errors clustered at hospital level

[a] $Proc^j_{i,h,t} = 1$ if the unscheduled high-risk newborn receives procedure measured in the outcome variable on the day of admission.

Table A13: Delays in Common Procedures

Congestion	(1) Hearing Tests	(2) Vaccinations	(3) Hearing Tests within 4 Days	(4) Vaccinations within 4 Days	(5) # Days before Hearing Tests	(6) # Days before Vaccinations
All	0.00502	0.00691	0.00406	0.00202	0.0438	0.0722
	(0.00893)	(0.00603)	(0.00942)	(0.00777)	(0.0825)	(0.0904)
N	23791	23791	23791	23791	14574	10174
Y-mean	0.613	0.428	0.433	0.341	3.093	2.047
Low	-0.00550	0.00859	0.00833	0.00222	-0.118	0.0667
	(0.0106)	(0.0122)	(0.0138)	(0.0109)	(0.136)	(0.151)
N	7015	7015	7015	7015	4313	3042
Y-mean	0.616	0.435	0.429	0.345	3.162	2.051
Medium	0.00696	-0.00501	-0.00246	-0.00290	0.124	-0.0275
	(0.0124)	(0.00965)	(0.0127)	(0.0110)	(0.116)	(0.124)
N	11790	11790	11790	11790	7258	5063
Y-mean	0.616	0.430	0.435	0.342	3.094	2.071
High	0.00734	0.0224	-0.00187	0.00540	0.158	0.381
	(0.0190)	(0.0271)	(0.0197)	(0.0246)	(0.161)	(0.240)
N	4986	4986	4986	4986	2990	2044
Y-mean	0.601	0.413	0.435	0.333	2.993	1.979

Standard errors in parentheses Standard errors clustered at hospital level

[a] Hearing tests and vaccinations are the top two procedure categories performed on newborns after birth.

Table A14: Treated - Control Differences:
Admissions on the Same Day or the Two Days Post to Unscheduled High-risk Admissions

Congestion	(1) C-section	(2) Birth Weight	(3) White	(4) Black	(5) Female	(6) Medicaid	(7) N
Panel A: Post - Control Difference							
Low	0.00619	-3.169	-0.00647	-0.00160	-0.0272	0.0146	6892
	(0.0186)	(14.51)	(0.0162)	(0.0173)	(0.0256)	(0.0159)	
Medium	0.0163	6.297	-0.00353	0.0145	0.0243	-0.0139	11774
	(0.0200)	(9.259)	(0.0101)	(0.0135)	(0.0176)	(0.0103)	
High	-0.0298	-8.511	-0.0361*	-0.00162	0.0359	0.0347*	5118
	(0.0210)	(12.63)	(0.0210)	(0.0242)	(0.0240)	(0.0202)	
Panel B: Same - Control Difference							
Low	-0.0420*	-1.533	-0.0175	0.00367	-0.0586**	0.0168	6731
	(0.0232)	(13.76)	(0.0164)	(0.0243)	(0.0251)	(0.0190)	
Medium	0.00118	11.80	-0.0238***	0.00121	0.0270	-0.0158	11255
	(0.0188)	(11.34)	(0.00847)	(0.0132)	(0.0235)	(0.0168)	
High	-0.0143	-8.322	-0.0364	-0.00608	-0.0646	0.0327	4783
	(0.0352)	(18.04)	(0.0247)	(0.0249)	(0.0383)	(0.0256)	

Standard errors in parentheses   Standard errors are clustered at hospital level

[a] Hospital-year, month, and day of week fixed effects are included in measuring treated–control differences.

Table A15: Treated - Control Differences:
Admissions on the Two Days Prior to C-section High-risk Admissions

Congestion	(1) C-section	(2) Birth Weight	(3) White	(4) Black	(5) Female	(6) Medicaid	(7) N
Low	-0.00819	-7.253	0.00317	0.0107	-0.00554	0.0217*	6597
	(0.0176)	(6.960)	(0.0109)	(0.00968)	(0.0161)	(0.0125)	
Medium	0.00146	-7.841	0.0100	-0.00217	-0.0301*	0.0161	10619
	(0.0143)	(7.777)	(0.0132)	(0.00965)	(0.0165)	(0.00994)	
High	-0.00561	-1.265	0.00472	-0.0153	0.00894	-0.00399	4417
	(0.0187)	(11.70)	(0.0238)	(0.0209)	(0.0223)	(0.0175)	

Standard errors in parentheses   Standard errors are clustered at hospital level

[a] Hospital-year, month, and day of week fixed effects are included in measuring treated–control differences.

Table A16: Spillover Effects among Newborns admitted Post to High-risk Admissions

Congestion	(1) Physician Tenure	(2) Experience with VLBW	(3) Avg. Len. of Stay (log)	(4) Avg. Total Charges (log)	(5) Avg. # of Procedures	(6) Avg. Res. Len. of Stay (log)	(7) Avg. Res. Total Charges (log)	(8) Avg. Res. # of Procedures
Low	0.448	0.00319	-0.00985	-0.00602	-0.0297	-0.000918	0.0182	-0.0107
	(0.459)	(0.00292)	(0.0264)	(0.0516)	(0.0585)	(0.00963)	(0.0291)	(0.0402)
Medium	-0.186	0.000549	-0.0163	-0.0122	-0.0159	-0.00400	-0.00323	-0.00300
	(0.420)	(0.00179)	(0.0132)	(0.0265)	(0.0269)	(0.00432)	(0.0118)	(0.0204)
High	0.199	-0.000483	-0.0405**	-0.0654*	-0.0719**	-0.00647	-0.00863	-0.0449*
	(0.494)	(0.00271)	(0.0190)	(0.0356)	(0.0342)	(0.00741)	(0.0195)	(0.0247)
N	23161	23657	23523	23523	23523	23523	23523	23523
Y-mean	17.15	0.0458	1.769	9.318	2.079	0.0614	0.200	0.209
	Len. of Stay (log)	Total Charges (log)	# of Procedures	In-hospital Mortality	Hospital Transfer	28-Day Readmission	1-Year Readmission	
Low	-0.00798	-0.00870	-0.0139	0.00398	-0.00244	-0.00830	-0.0139	
	(0.0265)	(0.0539)	(0.0820)	(0.00490)	(0.00505)	(0.00601)	(0.0121)	
Medium	0.0145	0.0184	0.0691	-0.00229	-0.00212	0.00205	0.00128	
	(0.0143)	(0.0328)	(0.0625)	(0.00200)	(0.00401)	(0.00463)	(0.0103)	
High	0.00618	-0.0194	0.0610	0.000657	-0.00395	0.00521	-0.00290	
	(0.0296)	(0.0474)	(0.113)	(0.00286)	(0.00598)	(0.00864)	(0.0127)	
N	23784	23784	23784	23784	23784	23784	23784	
Y-mean	1.875	9.435	2.682	0.00694	0.0128	0.0172	0.0690	

Standard errors in parentheses

Standard errors clustered at hospital level

Table A17: Spillover Effects among Newborns admitted on the Same Day of High-risk Admissions

Congestion	(1) Physician Tenure	(2) Experience with VLBW	(3) Avg. Len. of Stay (log)	(4) Avg. Total Charges (log)	(5) Avg. # of Procedures	(6) Avg. Res. Len. of Stay (log)	(7) Avg. Res. Total Charges (log)	(8) Avg. Res. # of Procedures
Low	-0.432	0.00893**	0.0407	0.0948*	0.0811*	0.00843	0.0414*	0.0269
	(0.639)	(0.00432)	(0.0258)	(0.0490)	(0.0465)	(0.00679)	(0.0236)	(0.0235)
Medium	0.549	0.00442	0.00672	0.00834	0.0354	0.00857	0.00984	0.0337
	(0.490)	(0.00636)	(0.0426)	(0.0816)	(0.101)	(0.0153)	(0.0391)	(0.0702)
High	-0.0731	-0.00105	0.0183	0.0567	0.0372	0.0124	0.0540*	0.0446
	(0.450)	(0.00380)	(0.0322)	(0.0658)	(0.0549)	(0.00835)	(0.0295)	(0.0296)
N	22169	22644	22520	22520	22520	22520	22520	22520
Y-mean	17.16	0.0456	1.770	9.318	2.081	0.0620	0.201	0.212
	Len. of Stay (log)	Total Charges (log)	# of Procedures	In-hospital Mortality	Hospital Transfer	28-Day Readmission	1-Year Readmission	
Low	0.0420	0.0675	0.246**	0.00925	0.00414	0.00000326	0.0109	
	(0.0329)	(0.0552)	(0.108)	(0.0118)	(0.00553)	(0.00493)	(0.0158)	
Medium	0.0354	0.0613	-0.0619	-0.00647***	0.00907	0.00461	0.0128	
	(0.0456)	(0.0835)	(0.125)	(0.00146)	(0.00631)	(0.00535)	(0.0164)	
High	0.0630	0.135*	0.296**	0.00841	0.00861	0.0168	0.0135	
	(0.0400)	(0.0742)	(0.140)	(0.00767)	(0.00802)	(0.0166)	(0.0188)	
N	22769	22769	22769	22769	22769	22769	22769	
Y-mean	1.875	9.435	2.681	0.00690	0.0134	0.0175	0.0697	

Standard errors in parentheses

Standard errors clustered at hospital level

Table A18: Spillover Effects from C-section High-risk Admissions

Congestion	(1) Physician Tenure	(2) Experience with VLBW	(3) Avg. Len. of Stay (log)	(4) Avg. Total Charges (log)	(5) Avg. # of Procedures	(6) Avg. Res. Len. of Stay (log)	(7) Avg. Res. Total Charges (log)	(8) Avg. Res. # of Procedures
Low	-0.112	0.00480	0.00906	0.0207	0.0411	0.00218	0.00952	0.0287
	(0.380)	(0.00404)	(0.0176)	(0.0372)	(0.0325)	(0.00457)	(0.0155)	(0.0173)
Medium	-0.198	-0.000167	0.0168	0.0391*	0.0304	0.00750	0.0230**	0.0220
	(0.328)	(0.00230)	(0.0138)	(0.0226)	(0.0300)	(0.00508)	(0.00850)	(0.0202)
High	0.0770	0.000961	0.00139	-0.00503	-0.00270	-0.0000184	-0.00387	-0.00425
	(0.435)	(0.00179)	(0.0200)	(0.0375)	(0.0379)	(0.00506)	(0.0129)	(0.0197)
N	21041	21511	21392	21392	21392	21392	21392	21392
Y-mean	17.19	0.0453	1.766	9.306	2.083	0.0610	0.199	0.210
	Len. of Stay (log)	Total Charges (log)	# of Procedures	Hospital Mortality	Hospital Transfer	28-Day Readmission	1-Year Readmission	
Low	-0.0229	-0.0298	0.0300	0.00528	-0.00528	-0.000859	-0.0167*	
	(0.0194)	(0.0365)	(0.0594)	(0.00487)	(0.00462)	(0.00430)	(0.00878)	
Medium	0.0145	0.0326*	0.00510	-0.00298	0.00152	0.00145	0.00621	
	(0.0116)	(0.0182)	(0.0395)	(0.00284)	(0.00239)	(0.00306)	(0.00557)	
High	0.0166	0.0592	-0.0112	0.00558	-0.00780**	0.00674	-0.00506	
	(0.0234)	(0.0394)	(0.0758)	(0.00333)	(0.00372)	(0.00435)	(0.00897)	
N	21632	21632	21632	21632	21632	21632	21632	
Y-mean	1.879	9.431	2.695	0.00707	0.0134	0.0170	0.0702	

Standard errors in parentheses

Standard errors clustered at hospital level

Table A19: LASSO Prediction Model for C-section High-risk Admissions

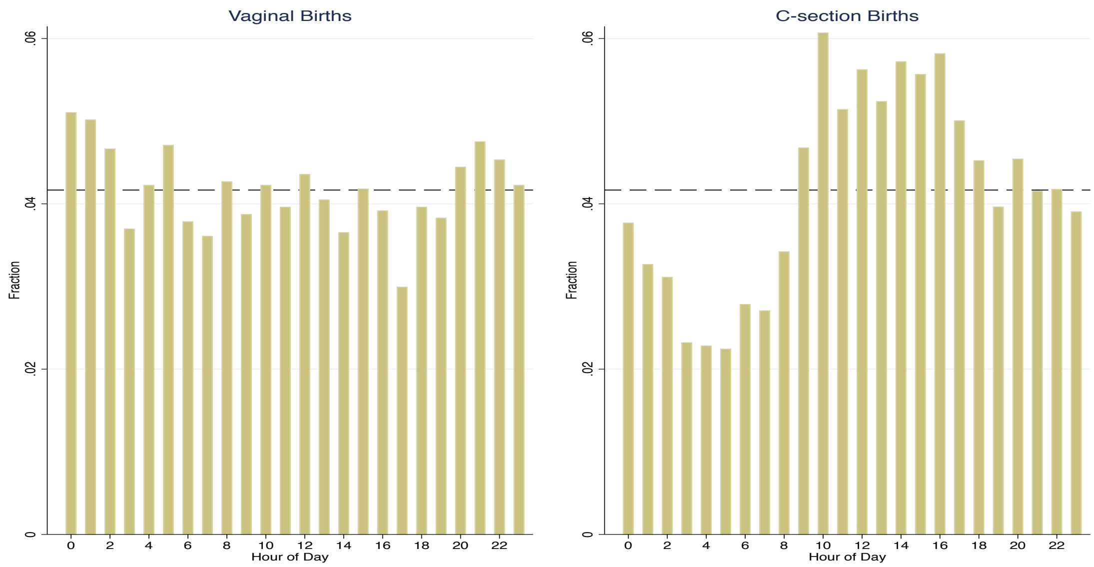
	LASSO[a]	Post-LASSO OLS
Physician Experience with VLBW (L2)	-0.0459	-0.100**
		(0.0473)
Physician Experience with LBW (L1)	0.0217	0.0488
		(0.0507)
Physician Avg. Res. # of Procedures (L1)	0.00155	0.00612
		(0.00950)
Total # of Attending Physicians (L1)	0.000366	0.000646
		(0.00134)
Total Newborn Admissions (L1)	0.000422	0.000605
		(0.000629)
Fraction of LBW (L3)	0.000936	0.00767
		(0.00915)
Fraction of VLBW (L2)	-0.0153	-0.0356**
		(0.0165)
Fraction of VLBW (L3)	0.0216	0.0363*
		(0.0206)
Fraction of White (L2)	0.00233	0.0106
		(0.00677)
Fraction of White (L3)	-0.00330	-0.0112
		(0.00716)
Fraction of Black (L3)	0.00247	0.00721
		(0.00436)
Month Dummy (Feb)	-0.00148	-0.00654
		(0.00394)
Month Dummy (May)	0.000810	0.00448
		(0.00273)
Month Dummy (Jun)	0.00137	0.00499
		(0.00475)
Month Dummy (Nov)	-0.00469	-0.00956**
		(0.00404)
Month Dummy (Dec)	-0.00921	-0.0139***
		(0.00420)
Day of Week Dummy (Wed)	0.00205	0.00471
		(0.00316)
Day of Week Dummy (Thu)	0.00127	0.00406
		(0.00371)
Day of Week Dummy (Sat)	-0.0104	-0.0139***
		(0.00334)
Day of Week Dummy (Sun)	-0.0115	-0.0144***
		(0.00351)
N	59311	59311

Standard errors clustered at hospital level are in parentheses

[a] Hospital-year fixed effects are included and not penalized. Parameter lambda is selected to minimize the mean-squared prediction error after cross-validation.
[b] L# indicates #-day lagged measure.

D Appendix Figures

Figure B1: Distribution of High-risk Admissions: Hour of Day



Note: The dash line is at $\frac{1}{24}$, showing an hypothetical uniform distribution.

Figure B2: Days between Unscheduled High-risk Admissions

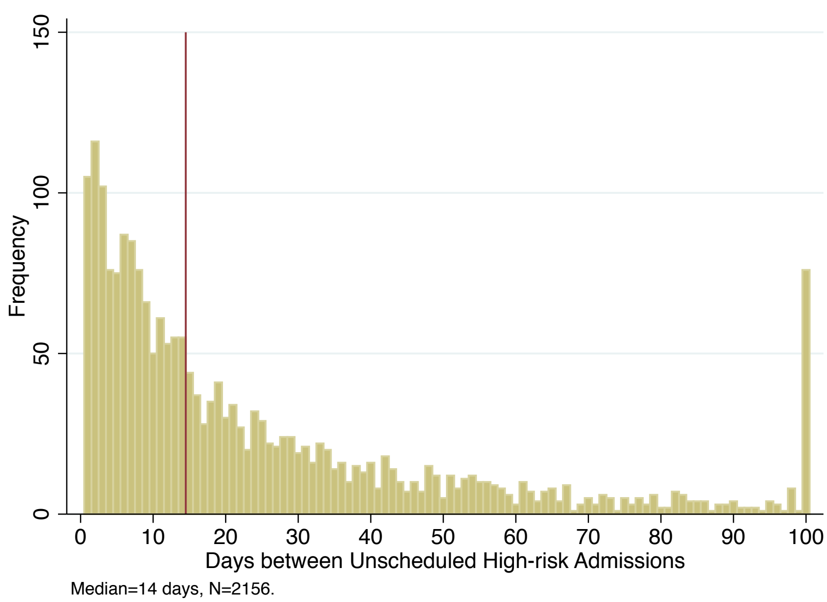


Figure B3: Variation in NICU Occupancy

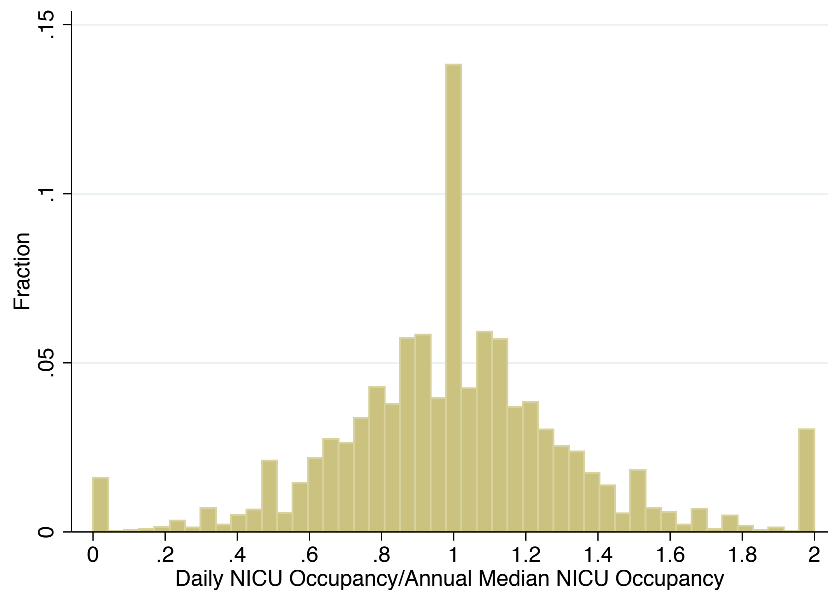


Figure B4: Average Total Charges by Birth Weight

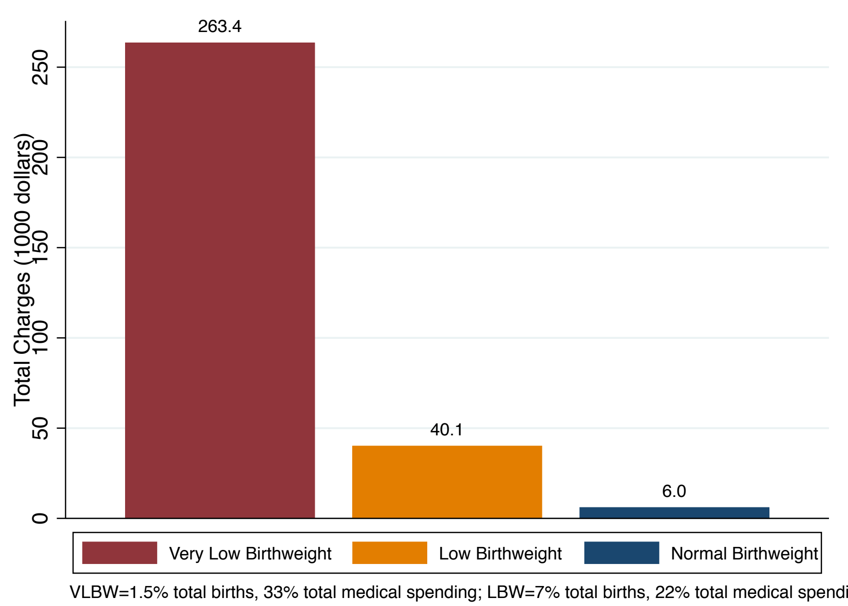


Figure B5: Probability of In-hospital Death by Birth Weight

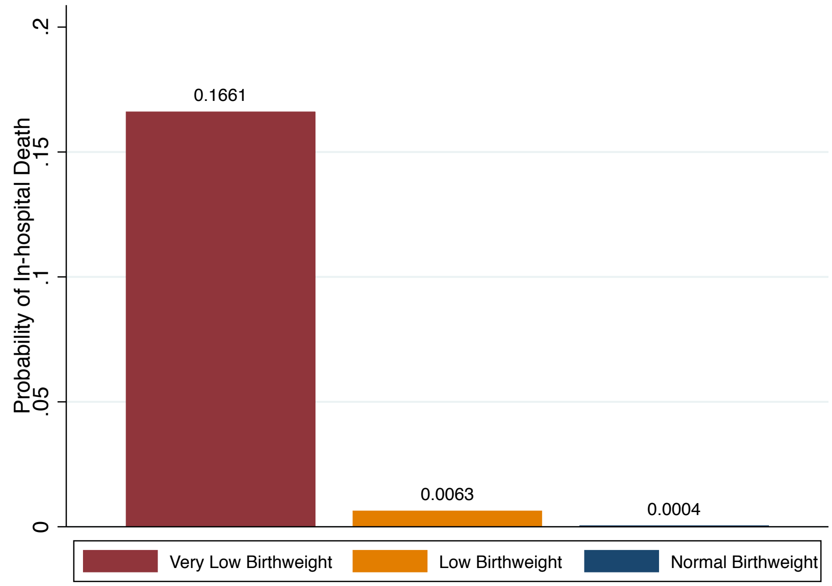


Figure B6: Distribution of Birth Weight

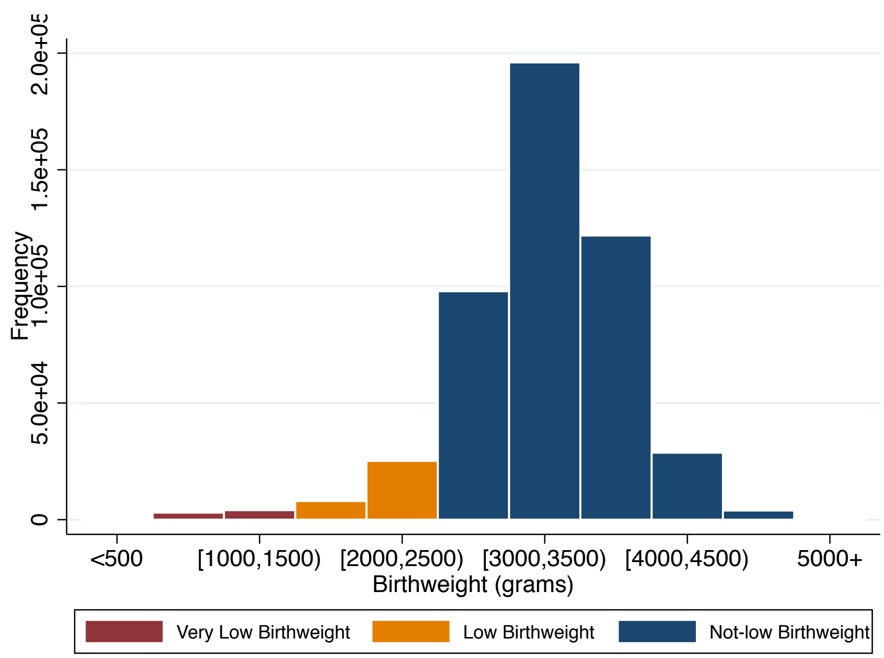
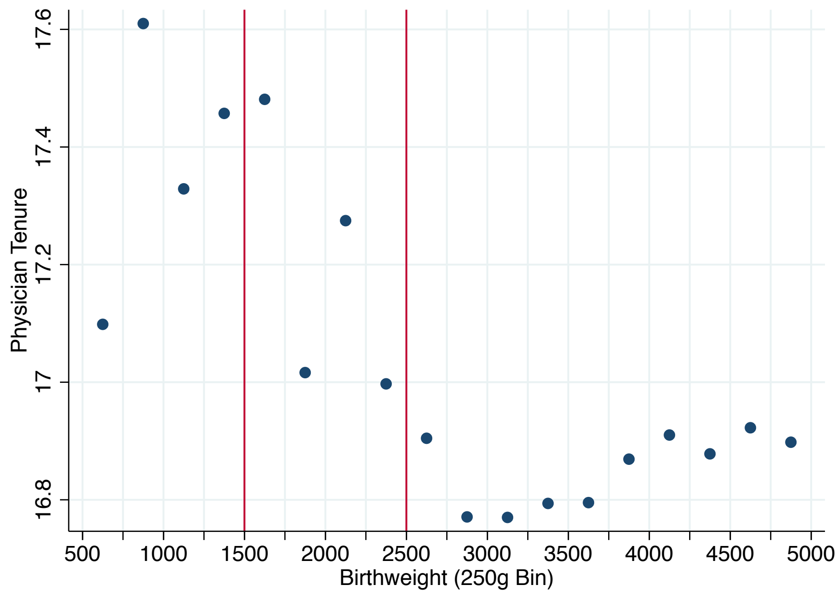
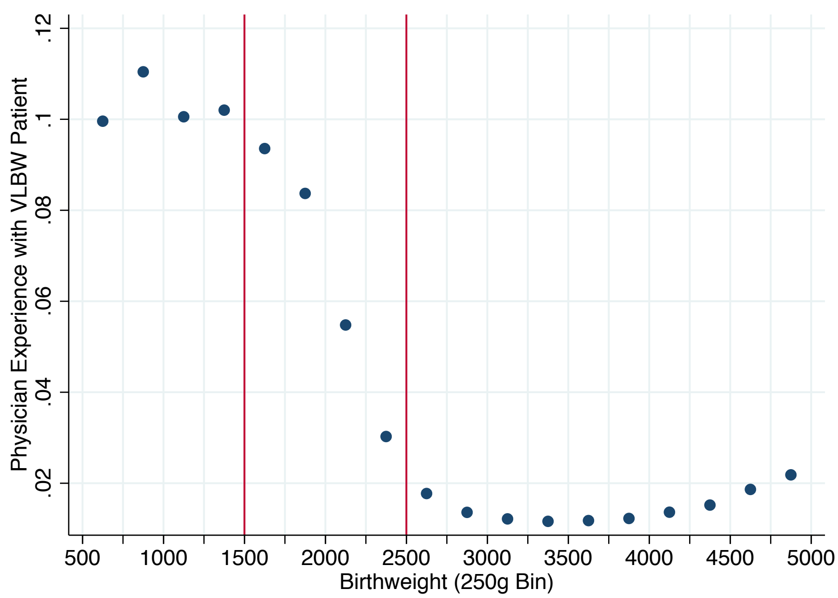


Figure B7: Physician Tenure by Birth Weight



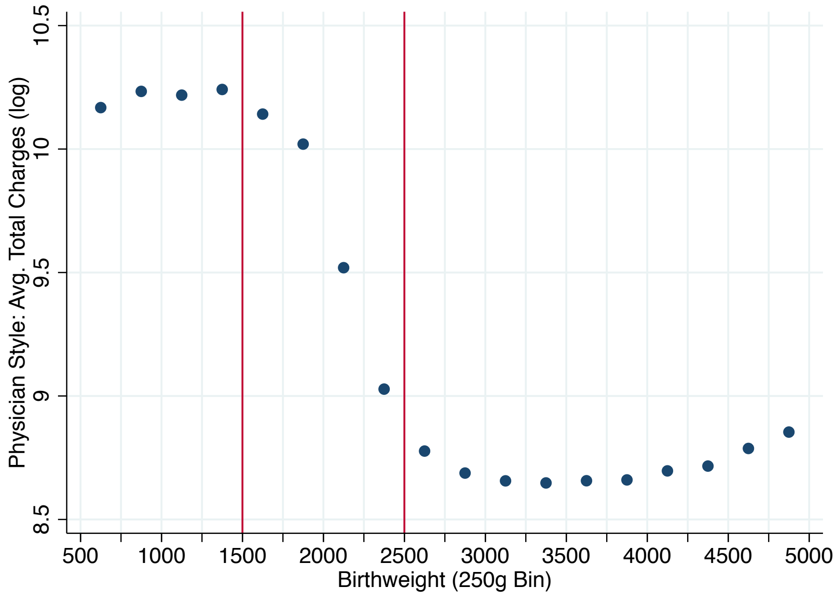
Newborns with birth weight below 500g or above 5000g are excluded.

Figure B8: Physician Experience with VLBW Patient by Birth Weight



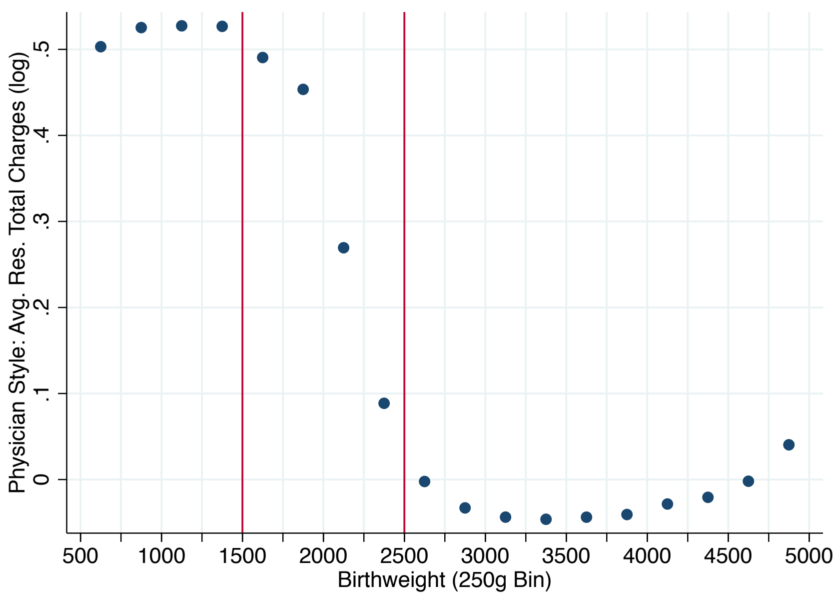
Newborns with birth weight below 500g or above 5000g are excluded.

Figure B9: Physician Style by Birth Weight - Baseline Measure



Newborns with birth weight below 500g or above 5000g are excluded.

Figure B10: Physician Style by Birth Weight - Residual Measure



Newborns with birth weight below 500g or above 5000g are excluded.

Figure B11: Daily Admissions and Discharges of NICU Patients

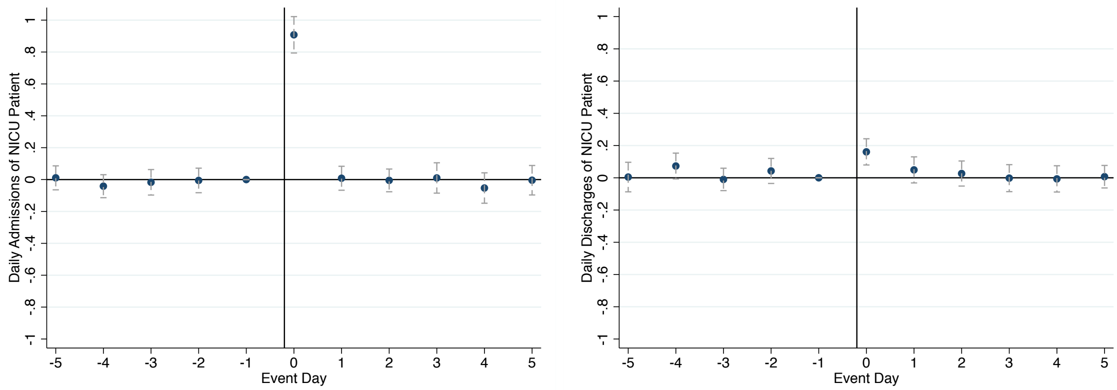
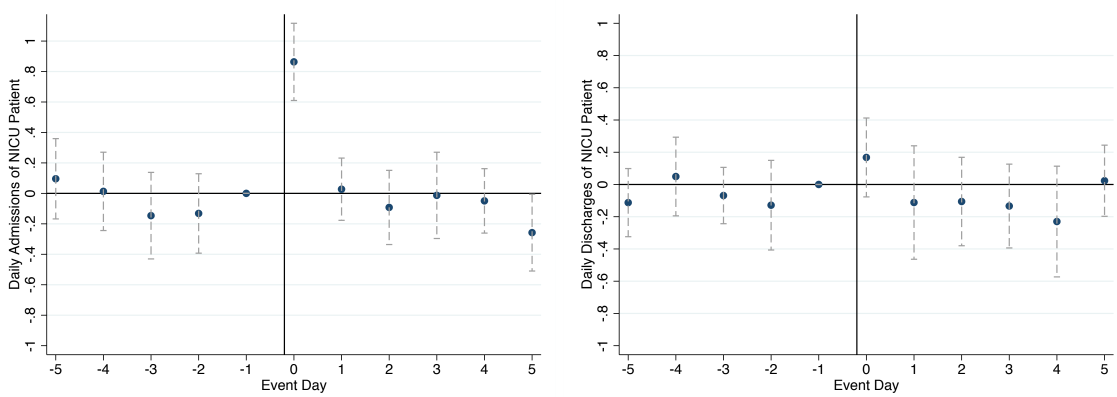
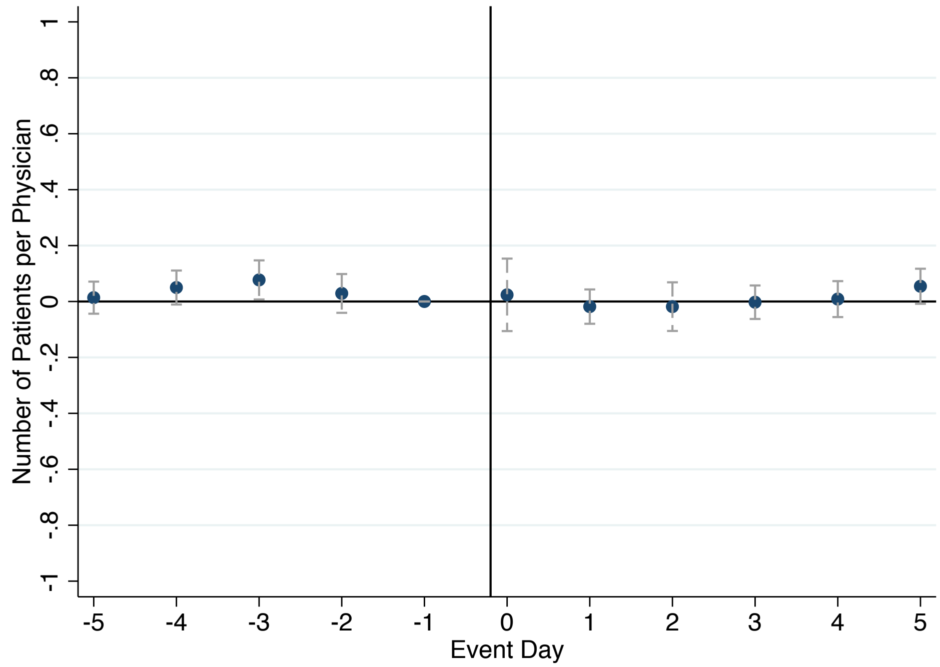


Figure B12: Daily Admissions and Discharges of NICU Patients
(high congestion on event day -1)



Graph generated by DD-event study to difference out the mean-reversion trend

Figure B13: Number of Patients Attended per Physician



The high-risk admission does not affect physician's workload

Figure B14: Number of Patients Attended by High-risk Newborn's Physician

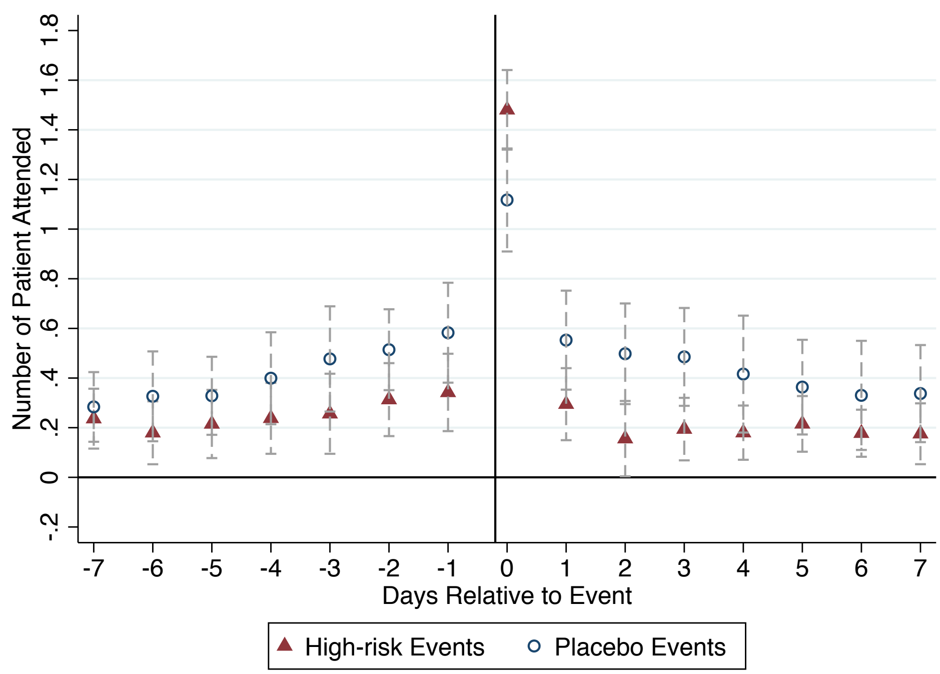


Figure B15: Number of Patients Attended by High-risk Newborn's Physician
(high congestion on event day -1)

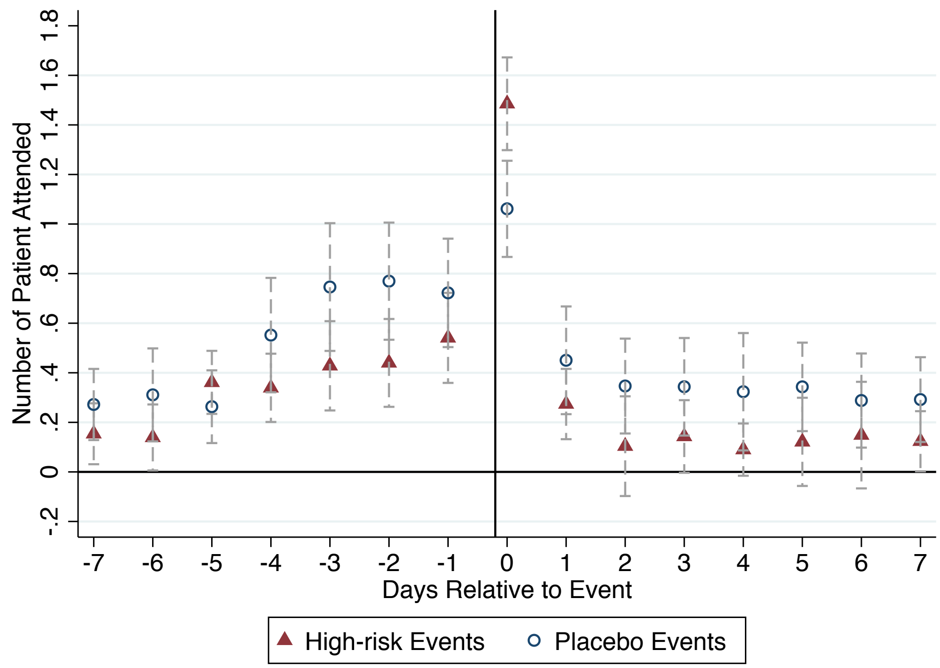


Figure B16: Number of Patients Attended in the Same Hospital
Prior to Days with Patient Admissions

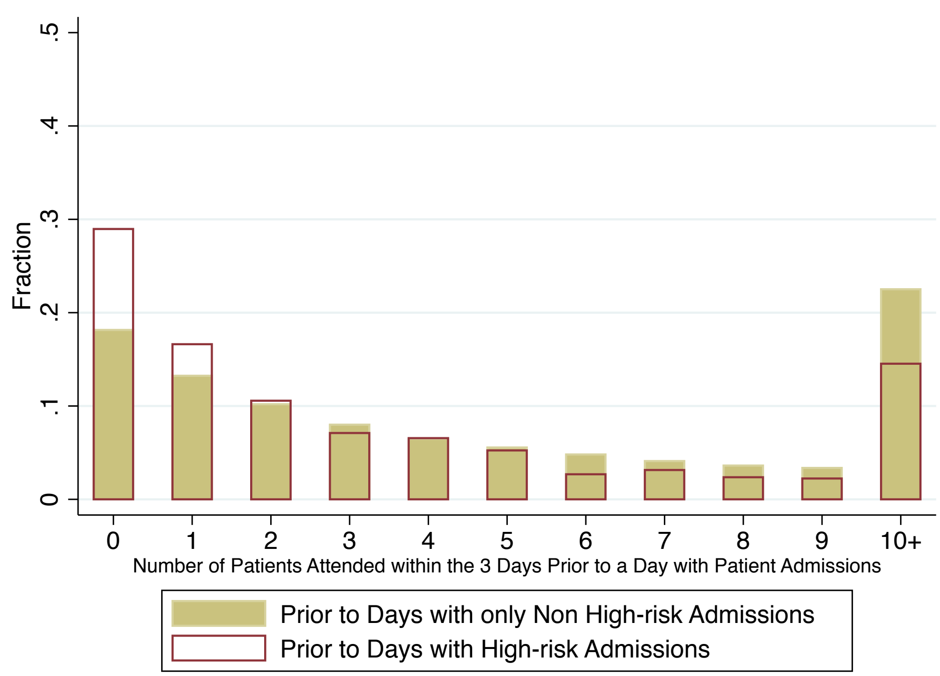
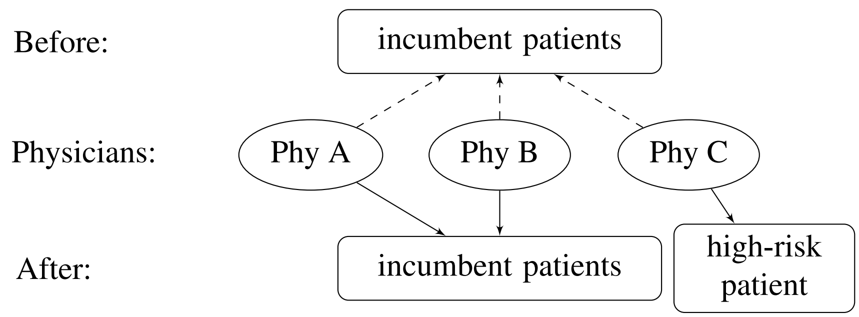
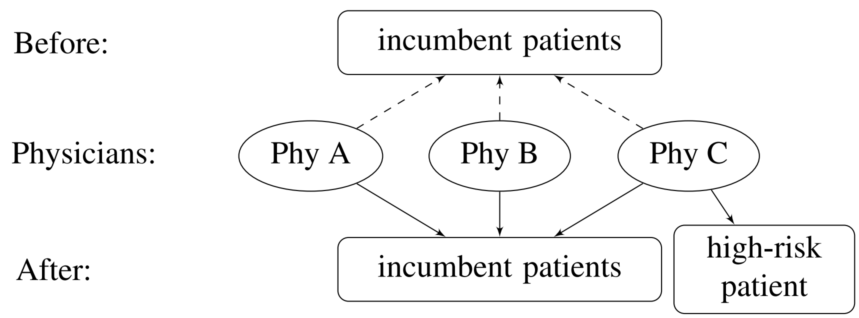


Figure B17: Hypothetical Hospitals' Responses

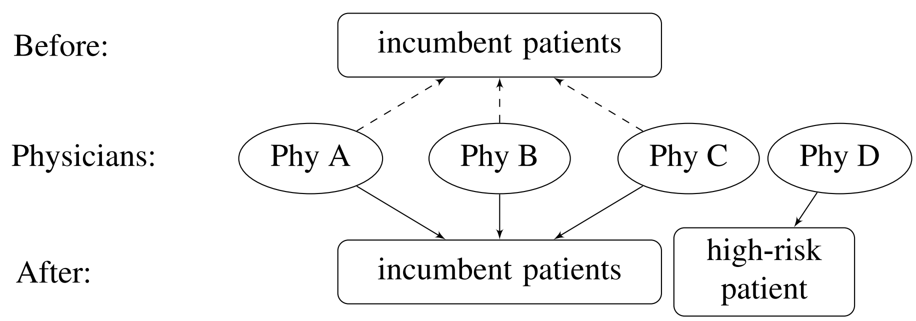
Scenario 1:



Scenario 2:



Scenario 3:



Scenario 4:

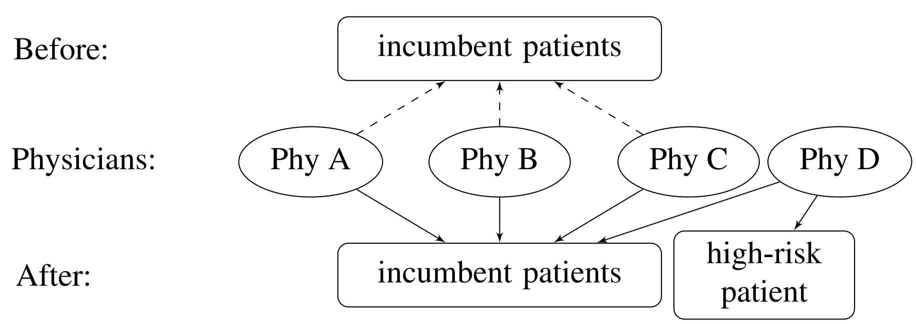
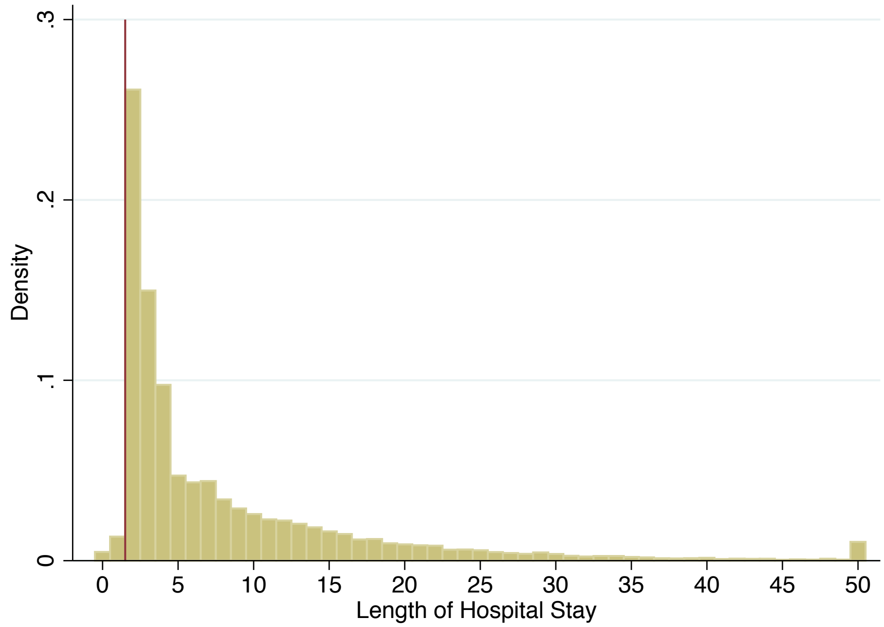
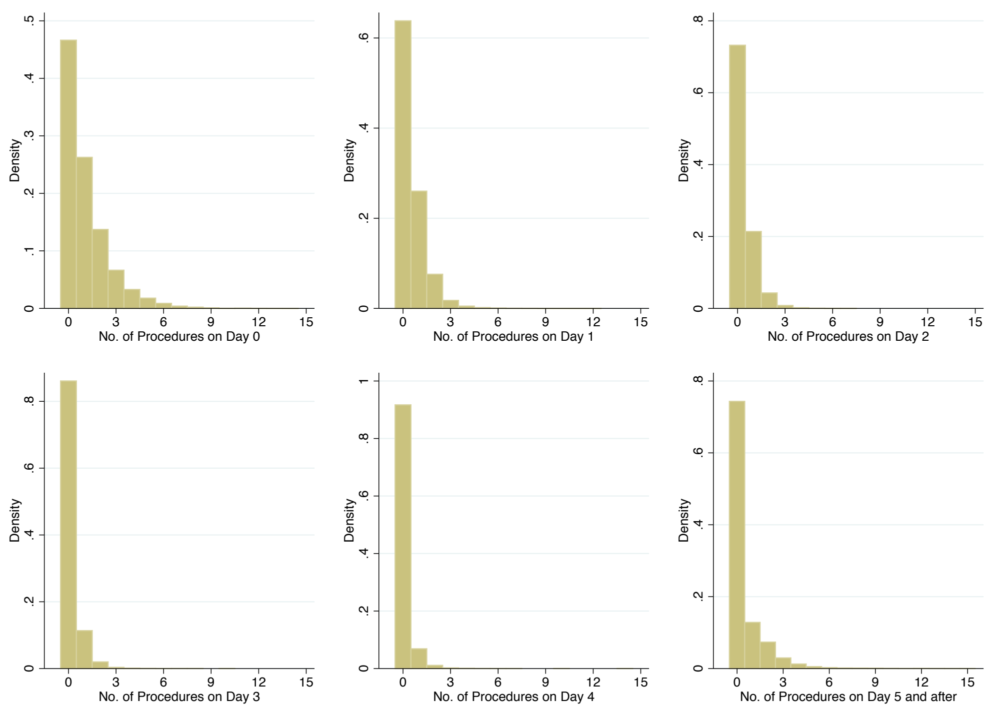


Figure B18: Distribution of Length of Hospital Stay



Sample: Singleton low birth weight newborns.

Figure B19: Distribution of Daily Procedures



Sample: Singleton low birth weight newborns.

Figure B20: Marginal Effect over Birth Weight

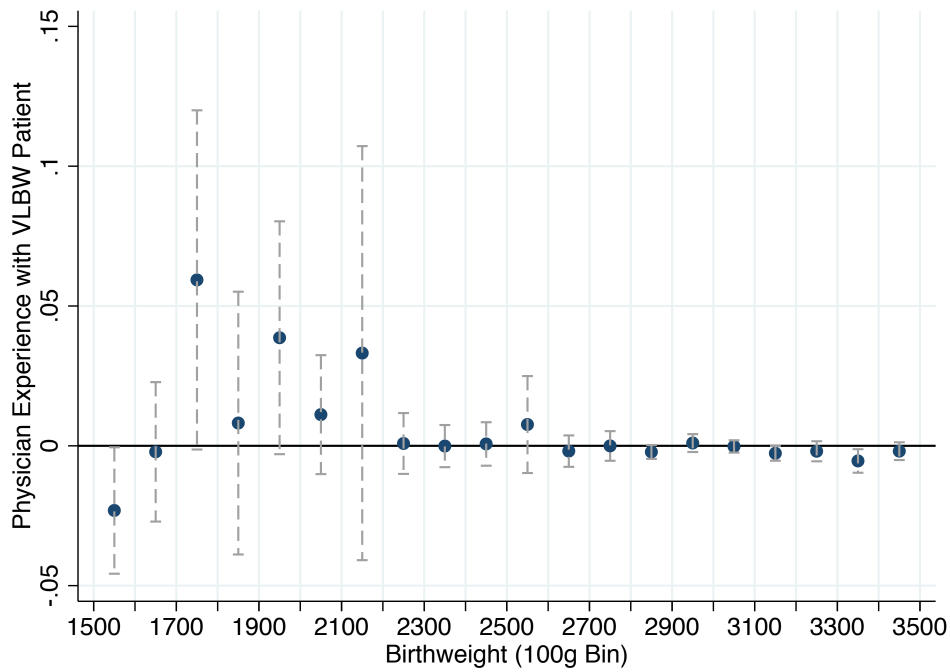
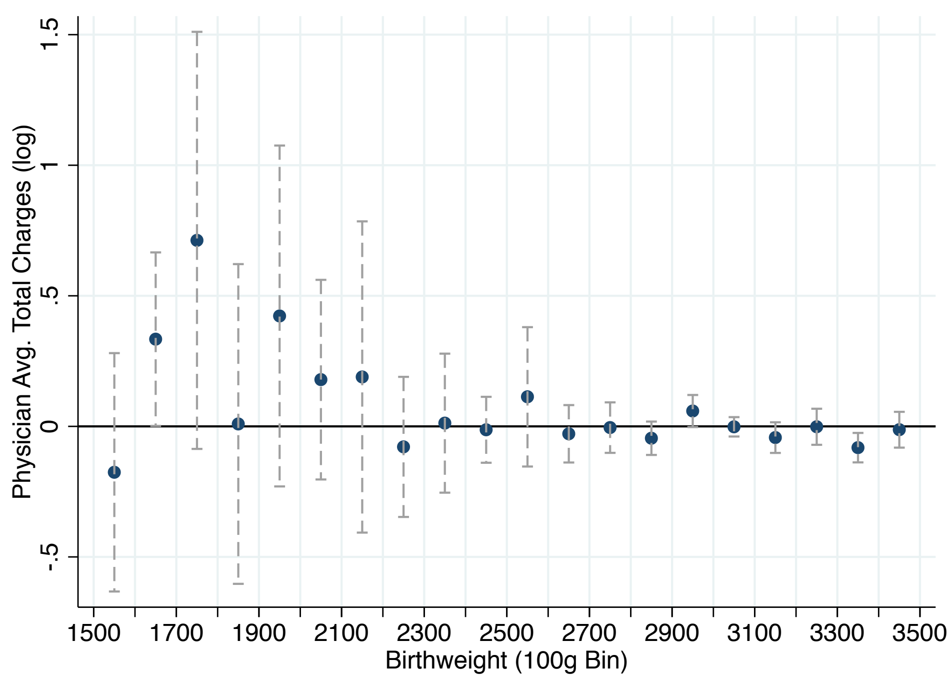
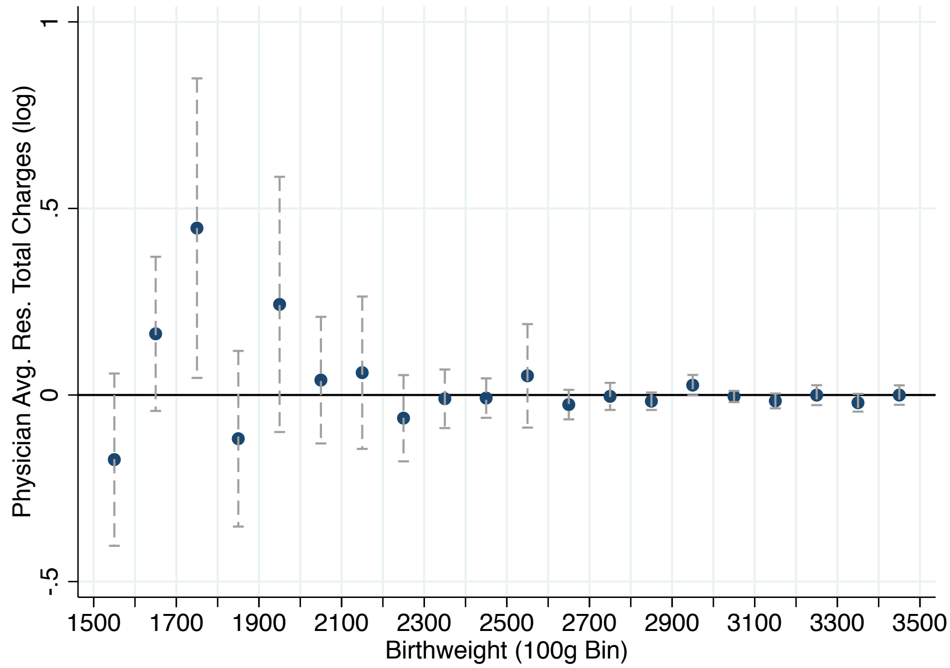


Figure B21: Marginal Effect over Birth Weight



Physician average length of stay and number of procedures show similar distribution.

Figure B22: Marginal Effect over Birth Weight



Physician average residual length of stay and number of procedures show similar distribution.

Figure B23: Marginal Effect over Birth Weight

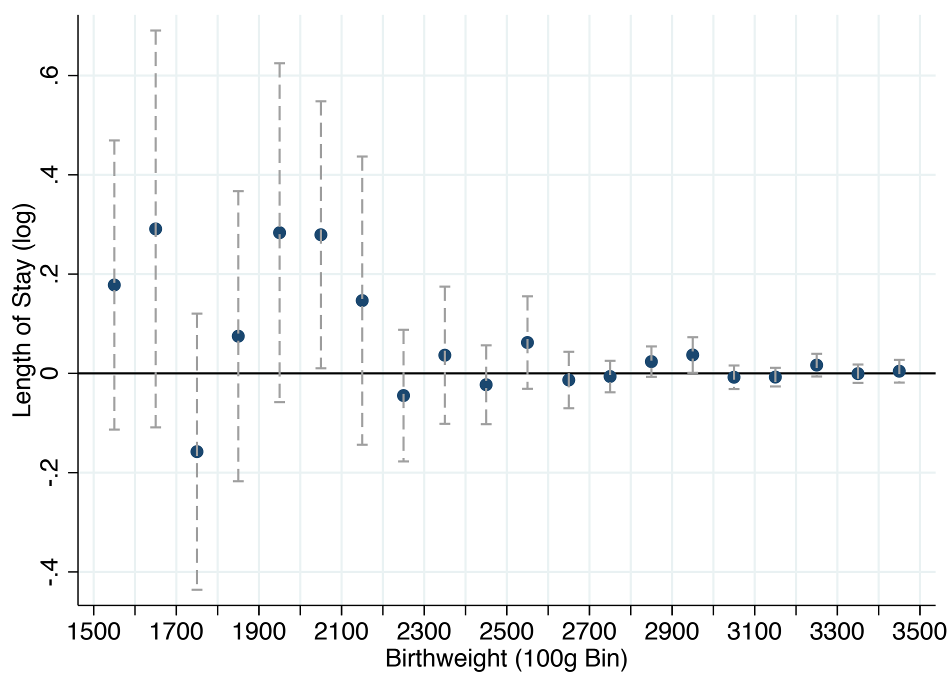


Figure B24: Marginal Effect over Birth Weight

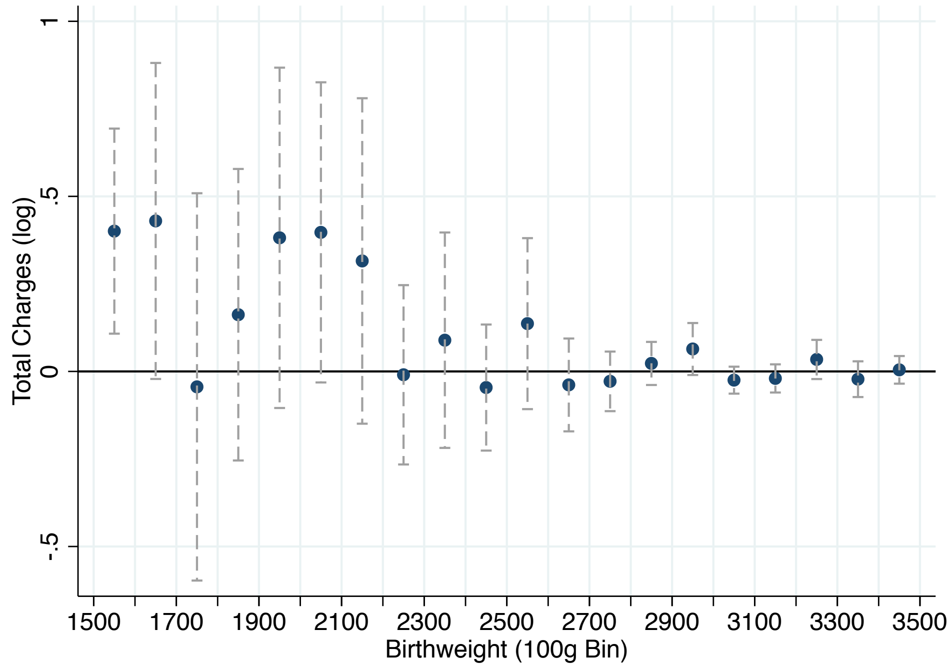


Figure B25: Marginal Effect over Birth Weight

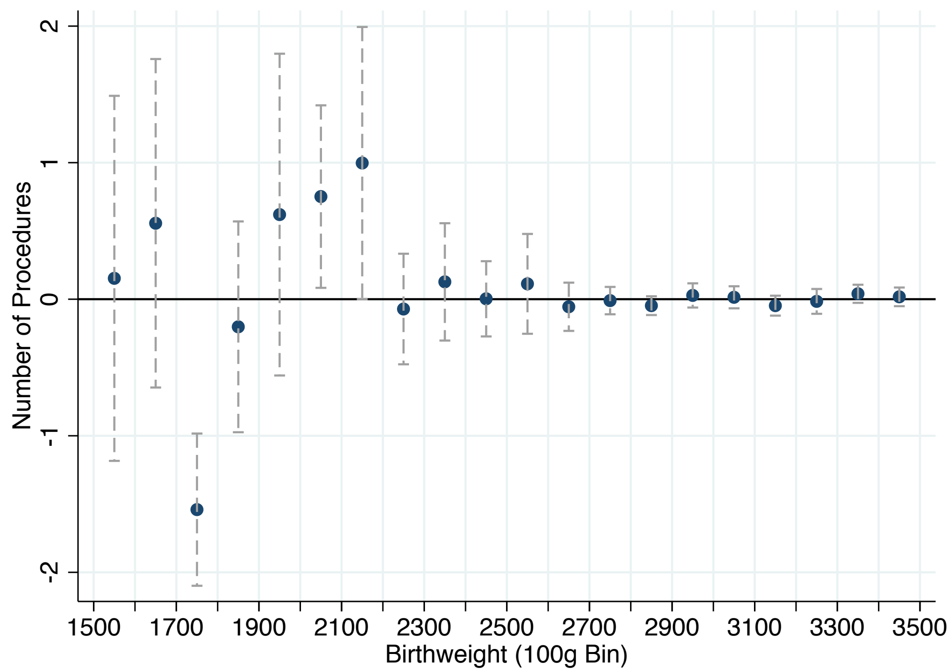
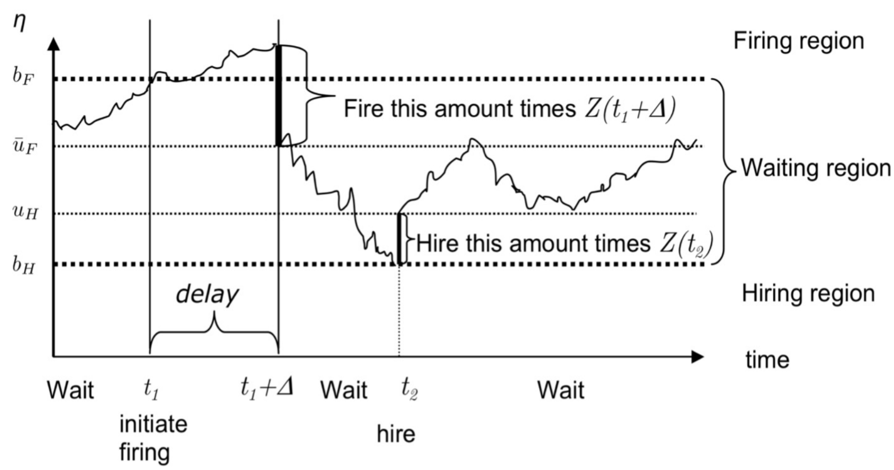


Figure B26: Solution Regions for Optimal Labor Control under Stochastic Demand



Source: Dai et al. (2015).

Dai et al. (2015) show that when faced with a stochastic demand, a firm does not constantly trace the optimal labor input level and only adjusts labor input when the labor-to-demand ratio η (the solid line) hits an upper or lower bound, i.e. b_F and b_H in the graph.