Essays in Health Economics

Essays in Health Economics

A dissertation presented

by

Johann Blauth

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

May 2014

Dissertation Advisor:
David M. Cutler

Author:
Johann Blauth

Essays in Health Economics

Abstract

This dissertation explores situational differences in physician behavior based on detailed electronic hospital records, shedding light on previously unobservable determinants of treatment decisions and processes.

I study how available operation room capacity affects operative decisions made by orthopedic surgeons in Chapter 1. These physicians have constrained access to operating room time, leading to exogenous variation of available capacity captured by the number of open surgery slots. I find that physicians are more likely to operate on a patient when they have more open surgery slots within a week of the initial encounter. This relationship is more pronounced for operative decisions allowing for substantial physician discretion. These results provide some evidence for how exogenous changes to available supply can have a significant impact on treatment decisions.

Chapter 2 — co-authored with David Ring and Mark S. Vrahas — exploits a similar measure of capacity to study the impact of operative delays on patient outcomes after hip fractures. Delaying the surgical treatment of these patients is frequently correlated with disadvantageous outcomes. Estimating the causal implications of this relationship is complicated by the influence of unobservable patient characteristics on both operative delays and outcomes. We address this issue by using the number of available surgery slots of the hospital's trauma surgeons at the time of patient arrival as an instrumental variable. This approach results in imprecise estimates of the causal consequences of operative delays despite a sufficiently strong first-stage relationship between operation room availability and operative delays.

In Chapter 3 — co-authored with Robert S. Huckman and David Ring — we study how the familiarity of surgical teams impacts procedure duration. Specifically, we analyze across what portion of a team's members familiarity must be developed to improve performance. We find that team familiarity — measured by the volume of previous operations performed by team members working together — reduces procedure duration by more to the extent that it is distributed across more members of a team. Broad familiarity shared by three or more team members improves team performance by more than twice as much as familiarity concentrated among just two team members.

Contents

Acknowledgments

I am deeply indebted to David M. Cutler for his continuous guidance and unwavering support. I will always remember his astute comments and precise questions, helping me to address weak points in my analysis and shaping the course of my research. I have greatly benefited from our discussions and the encouragement he provided so graciously.

Collaborating with Robert S. Huckman has been truly instructive and fruitful experience. He taught me how to use data to guide an analysis rather than blindly following an initial hypothesis. His feedback was always to the point and uplifting, contributing immensely to Chapter 3 of this dissertation.

Andrei Shleifer has been incredibly generous with his time and provided steadfast mentorship on all aspects of life. He helped me in keeping in mind the big picture in my research and in separating interesting from irrelevant questions. Without his resolute advice to finish timely, this dissertation would not have been completed at this point.

The research presented in this dissertation would not have been possible without the access provided by the Department of Orthopedics at Massachusetts General Hospital. I thank Jesse B. Jupiter for introducing me to the Hand and Upper Extremity Service and David Ring for his remarkable openness to new ideas and his invaluable support and trust. Working with Michiel G. Hageman made solving countless issues in accessing the data much less frustrating. Mark S. Vrahas was very gracious with his time to develop the ideas underlying Chapter 2 of this dissertation.

I received helpful feedback from Alberto Alesina, David Autor, Yvonne Braun, Raj Chetty, Tilman Dette, Nathaniel Hendren, Daniel Pollmann and numerous participants at the Labor and Public Finance Lunch at Harvard University and the Health Economics Seminar at the Department of Health Care Policy at Harvard Medical School.

I gratefully acknowledge financial support from the Department of Economics at Harvard University, the Douglas Dillon Fellowship Fund at Harvard University, and the German National Academic Foundation.

Introduction

This dissertation exploits detailed electronic hospital records bridging the gap between national databases — containing a limited number of variables on an enormous number of patients — and information collected through chart reviews, which typically comprises interesting covariates but is limited to a small number of patients. Here, I take advantage of data that would typically be collected through chart reviews but is available for a considerable number of patients.

Chapter 1 uses this unique source of data to measure a physician's access to operation room time and to determine whether this measure affects operative decisions. Chapter 2 exploits a similar measure of access to operation room time to instrument the impact of operative delays on patient outcomes after hip fractures. Both of these chapters are based on observing a physician's entire operation schedule over substantial periods of time. Meanwhile, Chapter 3 takes advantage of data tracking the composition of surgical teams to distinguish how procedure duration changes with familiarity concentrated among few members and familiarity shared by many members of a surgical team.

All of these chapters demonstrate situational differences in physician decisions and performance, helping us to understand further determinants of the abundant variation in treatment choices. Most importantly, however, situational variation — to the extent that it is exogenous to patient characteristics — constitutes a promising instrumental variable, enabling us to estimate the causal implications of physician behavior and treatment choices for patient outcomes. Chapter 2 proceeds accordingly, using situational variation in operation room availability as an instrumental variable for operative delay.

Physician Capacity and Operative Decisions

Clearly, treatment decisions should be influenced by patient and physician characteristics. Besides these health factors, non-health factors such as financial incentives and available capacity are frequently suspected to influence patient and physician behavior. So far, there is very limited empirical evidence on the role of physician capacity, presumably because identifying and tracking exogenous variation in physicians' time and access to resources is challenging.

Chapter 1 explores how operative decisions — among the most meaningful choices made by patients and physicians both in qualitative and financial terms — are affected by the available capacity of orthopedic surgeons. In orthopedic surgery, many conditions may be treated both non-operatively and surgically, leaving substantial discretion to the patient and the physician. I observe short-term fluctuations in orthopedic surgeons' access to operation room time and use the resulting variation to estimate the impact of open surgery slots on subsequent operative decisions. For identification, the variation in physician capacity must be unrelated to patient and physician characteristics. I focus on the initial encounters between patients and physicians to ensure that physician capacity is exogenous to patient characteristics. Meanwhile, I account for physician fixed effects, comparing the operative decisions made by the same physician over time.

I find that orthopedic surgeons are more likely to operate on a patient when they have more open surgery slots within a week of the initial encounter. A one standard deviation increase in physician capacity leads to an operation rate that is between 1.7 and 3.7 percent higher relative to the corresponding average operation rate over a monthly, quarterly, and yearly horizon. These differences are considerable given that I consider the entire spectrum of orthopedic conditions not requiring immediate treatment. Interestingly, physician capacity has a more pronounced impact on operative decisions allowing for substantial physician discretion. These results provide some evidence for how exogenous changes to available supply can have a significant impact on treatment decisions made by both patients and physicians.

Operative Delays and Hip Fracture Outcomes: Operation Room Availability as Instrumental Variable

Documenting situational variation in physician behavior helps us to further comprehend the enormous variation in treatment decisions we observe across many fields of medicine. Nonetheless, tracking situational variation may be even more relevant for the evaluation of the causal consequences of treatment processes and decisions. To the extent that situational variation is exogenous to patient characteristics, it may be used to instrument a wide range of treatment characteristics — which are typically endogenous to unobservable patient characteristics — when analyzing their impact on patient outcomes.

Chapter 2 — co-authored with David Ring and Mark S. Vrahas — pursues this strategy to address a long-standing issue in clinical research. Across many conditions in orthopedic surgery, adverse patient outcomes are positively correlated with operative delays. Frequently, severe cases require more time to be stabilized before surgery, leading to a spurious correlation between operative delays and patient outcomes. Clearly, estimating the causal implications of this relationship is complicated by the influence of unobservable patient characteristics on both operative delays and outcomes. There is still no consensus on how badly operative delays affect patient outcomes.

We focus on patients reporting to the emergency room with hip fractures, one of the most frequent conditions in orthopedic surgery. These patients are typically operated on by attending trauma surgeons. We exploit variation in these surgeons' number of open surgery slots at the time of patient arrival as instrumental variable for operative delay, arguing that operation room availability is unlikely to be related to patient characteristics conditional on the time of patient arrival. We document a small but significant relationship between operation room availability and operative delay, suggesting that increasing operation room availability by one standard deviation reduces operative delay by almost three hours, or about 6.8 percent of the average operative delay. Despite this first-stage relationship, the estimates of the causal consequences of operative delays on patient outcomes are imprecise.

Concentrated and Broad Team Familiarity: Evidence from Orthopedic Surgery

In Chapter 3 — co-authored with Robert S. Huckman and David Ring — we study how the familiarity of surgical teams impacts procedure duration of orthopedic operations. Specifically, we analyze across what portion of a team's members familiarity must be developed to improve performance. That is, must all members of a team work together or can proportional benefit be derived from just a subset of team members building experience with each other?

Typically, a team's familiarity is measured by the number of prior pairwise collaborations among its members. We develop a new measure of team familiarity, distinguishing between prior operations developing familiarity concentrated among two team members and past procedures building broad familiarity across three or more team members. We find that operations building broad familiarity reduce procedure duration by more than twice as much as operations developing concentrated familiarity. Our results imply that the benefits of broad familiarity can be replicated by a proportional amount of concentrated familiarity, suggesting that team familiarity can be build in pieces.

The effects of both concentrated and broad familiarity decay rapidly in the setting we study. A prior operation performed by two and three or more team members reduces the expected duration of the focal procedure by 2.2 and 5.3 percent, respectively, if it is performed on the same day prior to the focal operation. By contrast, an operation performed within 1–10 days before the focal procedure reduces its average duration by 0.1 percent if it develops concentrated familiarity and by 0.2 percent if it develops broad familiarity. We do not find effects of concentrated and broad familiarity developed more than 10 days before the focal procedure.

Chapter 1

Physician Capacity and Operative Decisions

1.1 Introduction

Physicians have long been suspected to make treatment choices not only based on patient needs. In his seminal contribution, Arrow (1963) points specifically to the role of financial incentives and available time as determinants of physician decisions. An enormous body of literature explores the implications of financial incentives in medical care, suggesting that physicians are considering monetary objectives in their decisions.[1] By contrast, there is very limited evidence on the impact of available capacity on physician decisions.

Despite the absence of empirical evidence, capacity has been linked to the enormous variation in treatment decisions witnessed across many areas of medicine. Fisher et al. (2003) document large variations in regional Medicare spending that are unrelated to patient outcomes, leading to an enormous debate about the validity and implications of their findings (Chandra and Staiger 2007, Doyle 2011). If treatment decisions varied with physician capacity, high-spending regions may be spending more because their physicians have more capacity. Physician capacity may thus be an important piece to solving this puzzle, especially in the context of "flat-of-the-curve" medicine practiced even when the marginal returns to treatment are fairly small (Enthoven 1980, Fuchs 2004).

[1] See for instance Cutler (1995), Dafny (2005), and Clemens and Gottlieb (2014).

Analyzing the impact of physician time on treatment choices is complicated by multiple issues. First and foremost, data capturing physician capacity is rare, leading researchers to rely on state-level variation in capacity relative to patient demand.[2] Second, disentangling time from other determinants of physician decisions constitutes a formidable challenge, because several important determinants — such as patient characteristics and physician skills — are not perfectly observable. Consequently, variation in physician capacity needs to be plausibly exogenous to any unobservable determinants of physician decisions to allow for convincing identification.

In this chapter, I study whether physician capacity affects the operative decisions made by orthopedic surgeons. My analysis takes advantage of short-term fluctuations in physician time stemming from institutional constraints in orthopedic surgery. In this area of medicine, physicians have to schedule operation room time way in advance. By contrast, individual operations are scheduled weeks or even days before the operation. This *modus operandi* is feasible because many orthopedic procedures are elective — and thus not time sensitive — and leads to occasional excess demand relative to the supply of surgery slots, resulting primarily from two sources. First, scheduled operation room capacity varies depending on the physician's presence in the hospital and typically decreases before a vacation, conference, or other external obligation. Second, the extent to which scheduled operation room capacity is already booked up varies with the number and severity of previously seen patients.

Operative decisions constitute the primary outcome measure of this study. Besides their obvious relevance to patients, operative decisions are an advantageous outcome because the need for surgery is rarely perfectly clear. Many orthopedic conditions can be addressed using either surgical or non-operative treatments, resulting in enormous physician discretion in whether or not to operate. As comparable outcomes for patient treated surgically and non-operatively are not collected in my setting, I am unable to analyze the relationship between operative decisions and health outcomes. I consider treatment costs as secondary outcome measure.

[2]For instance, Gruber and Owings (1996) exploit variation in fertility rates across states.

Of course, operative decisions vary with patient and physician characteristics that are unobservable. I rely on two strategies to address these identification issues. First, I concentrate on patients seeing an orthopedic surgeon for the first time during an office encounter scheduled in advance with the physician's secretary. I argue that unobservable patient characteristics are unlikely to be endogenous to physician capacity in this setting. Second, I focus on analyzing operative decisions of the same physician, taking advantage of short-term fluctuations in physician capacity. To the extent that unobservable physician characteristics do not fluctuate within short-term periods, this strategy ensures that results are not driven by differences in physician expertise.

I find that short-term physician capacity at the time of the initial encounter between a patient and an orthopedic surgeon significantly increases subsequent operation rates. Relative to the mean operation rate of 8.6, 16.9, and 22.7 percent within 30, 90, and 360 days, respectively, a one standard deviation increase in physician capacity raises the operation rate by 3.7 percent within a month, 3.2 percent within a quarter, and 1.7 percent within a year. To explore whether these effects are nonlinear, I assign patients to quintiles based on the physician's capacity during their encounter. Within 90 days, operations rates are 7.1 percent higher in the top quintile relative to the bottom quintile. Furthermore, I demonstrate that the impact of physician capacity on operation rates varies systematically with medical discretion. Capacity has almost no impact on the treatment of patients suffering from diagnoses allowing for limited physician discretion such as carpal tunnel syndrome. By contrast, conditions that can be treated both surgically and non-operatively — for example hand osteoarthritis — feature operation rates fluctuating substantially with available capacity.

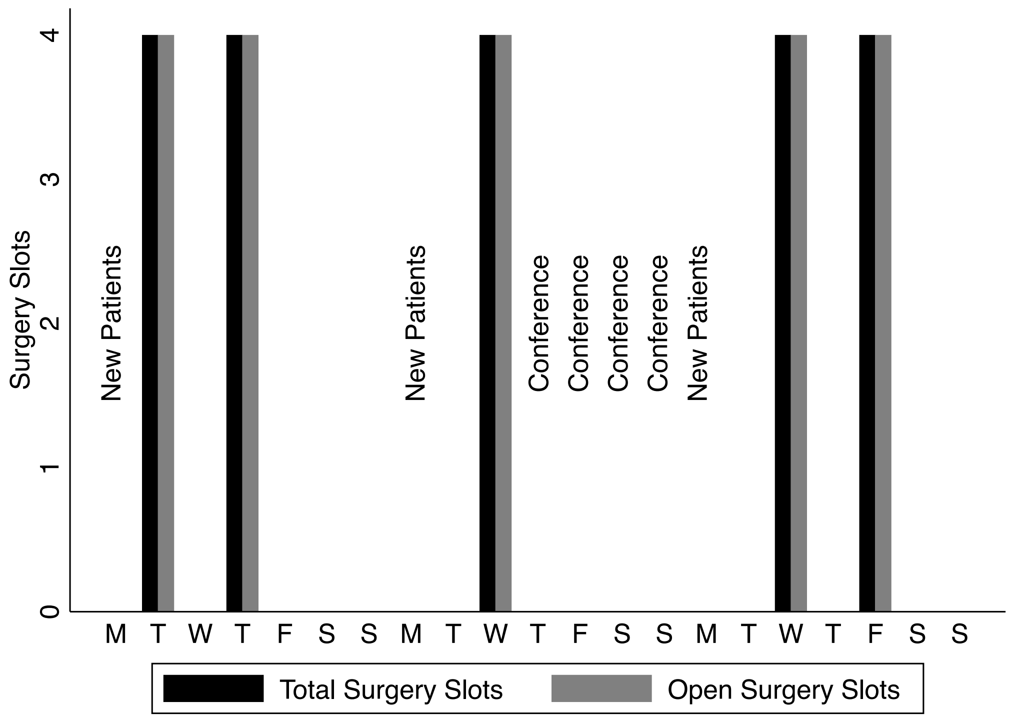
In general, treatment choices are determined by demand, supply, and situational factors (Chandra et al. 2011). As physician time is related to supply and situational factors, this chapter borrows from both strands of the literature. A large body of research examines whether physicians induce demand by changing patients' desired treatments (McClellan 2011). Initial studies demonstrate a positive relationship between the intensity of service provision and the number of physicians relative to the population (Fuchs 1978, Cromwell and Mitchell 1986), but suffer from identification issues (Dranove and Wehner 1994). Gruber and Owings (1996) address these challenges by taking advantage of an exogenous demand shocks to disentangle supply and demand factors. I use a similar strategy while exploiting changes in supply and relying on much more precisely identified short-term variation rather than state-level changes in the occurrence of medical conditions.

I follow the situational literature — such as Choudhry et al. (2006) — in exploiting physician-level rather than regional variation in operation rates. This strategy ensures that my analysis is not biased by physician characteristics, addressing one of the most pressing concerns raised in the analysis of the regional variation in operation rates (Glover 1938, Wennberg and Gittelsohn 1973, Fisher et al. 2003). These studies claim that patient and physician characteristics are balanced across regions, an assumption that has been questioned in recent work (Chandra and Staiger 2007, Doyle 2011). My study addresses both of these issues based on physician-level variation. Besides controlling for physician characteristics, using physician-level variation ensures that physician capacity is exogenous to patient characteristics.

This chapter proceeds as follows. Section 1.2 further explains the variation in physician capacity I observe based on the example of a typical schedule of an orthopedic surgeon. Section 1.3 describes the setting and data, emphasizing the identification and utilization of variation in physician time that is plausibly exogenous to patient characteristics. In Section 1.4, I outline the empirical strategy and discuss the estimation framework underlying this chapter. Section 1.5 presents the corresponding results. Finally, Section 1.6 concludes.

1.2 Capacity Constraints in Orthopedic Surgery

In the setting studied in this chapter, orthopedic surgeons have to schedule operation room time far in advance. Typically, physicians do not know which patients they are going to operate when they request operation room time. Thus, they have to estimate patient demand based on their experience. According to anecdotal evidence, they also

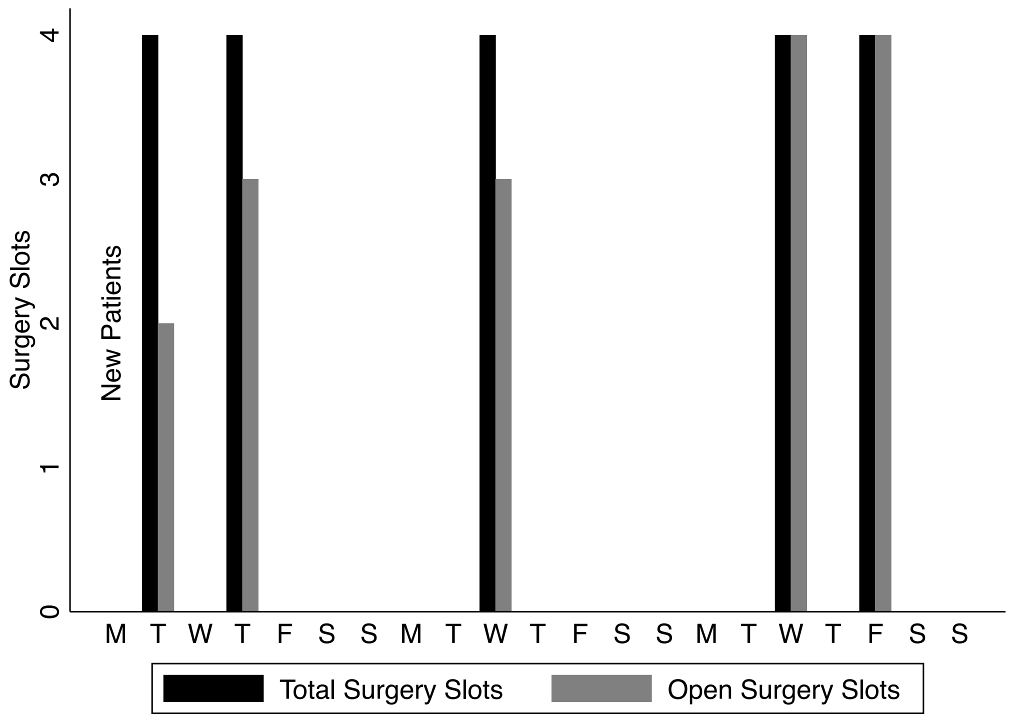


Notes: This figure shows the hypothetical schedule of an orthopedic surgeon observed when the surgeon schedules operation room time far in advance of the observation period.

Figure 1.1: *Typical Schedule of Orthopedic Surgeon*

consider obligations that are scheduled long in advance, including meetings, conferences, and vacations. Orthopedic surgeons are able to handle their operation schedule this way because many orthopedic surgeries are elective and do not need to be performed urgently.

Figure 1.1 illustrates a hypothetical schedule of an orthopedic surgeon over a three-week period at the time of requesting operation room time. Here, we see the physician request two operation days per week during the first and third week. During the second week, the surgeon plans to attend a conference, and thus plans just one operation day. While the surgeon does not know how many surgeries he will actually perform during each operation day — this number depends very much on the type of procedure as well as patient characteristics — he expects to perform about four procedures per day. Figure 1.1 also shows that the physician is seeing new patients in his office during each Monday of the

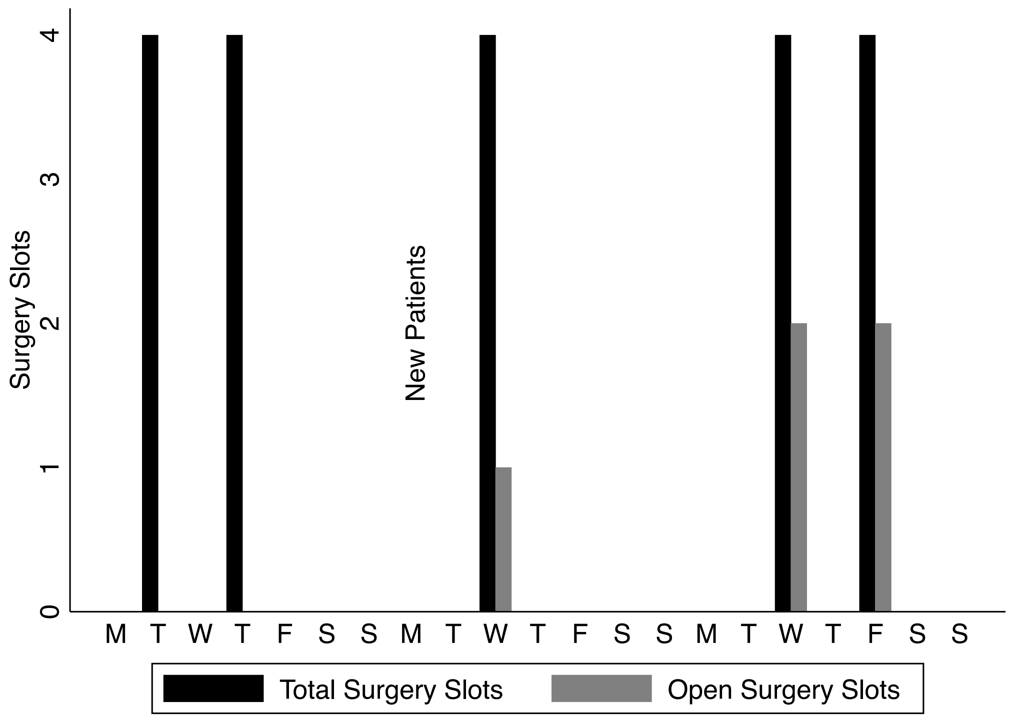


Notes: This figure shows the hypothetical schedule of an orthopedic surgeon observed on the first Monday of the observation period.

Figure 1.2: *Typical Schedule of Orthopedic Surgeon — First Monday*

observation period. These patients' subsequent operative decisions are the central outcome variable analyzed in this paper.

In Figure 1.2, we observe the physician's schedule at the start of the first Monday of the observation period. At this point, the surgeon has already assigned some patients surgery slots during the first three operation days. Thus, he has five operation slots available within a week of the seeing patients during the first Monday. Another three surgery slots are available within two weeks. In my setting, such a situation would reflect ample spare capacity. Figure 1.3 visualizes the surgeon's schedule at the start of the second Monday of the observation period. Now, the physician has only one operation slot available within a week of seeing patients during the second Monday, and four more open surgery slots within two weeks. This situation reflects fairly low short-term capacity.



Notes: This figure shows the hypothetical schedule of an orthopedic surgeon observed on the second Monday of the observation period.

Figure 1.3: *Typical Schedule of Orthopedic Surgeon — Second Monday*

In this chapter, I analyze whether patients initially seen on the first Monday are more likely to be operated than patients first seen during the second Monday. Put differently, I estimate whether the physician's open surgery slots within a week of the initial encounter with a patient affect the patient's subsequent operative decision. As shown in Figures 1.2 and 1.3, the variation in open surgery slots stems from two sources. The number of scheduled operation days within a week of the initial meeting with a patient is twice as high during the first Monday compared to the second Monday. In addition, the only operation day available within a week of the second Monday is far more booked than either of the operations days during the first week. It is important to note that both sources of variation need to be considered to obtain an accurate measure of physician capacity. I discuss this phenomenon in more detail in the next section.

1.3 Setting and Data

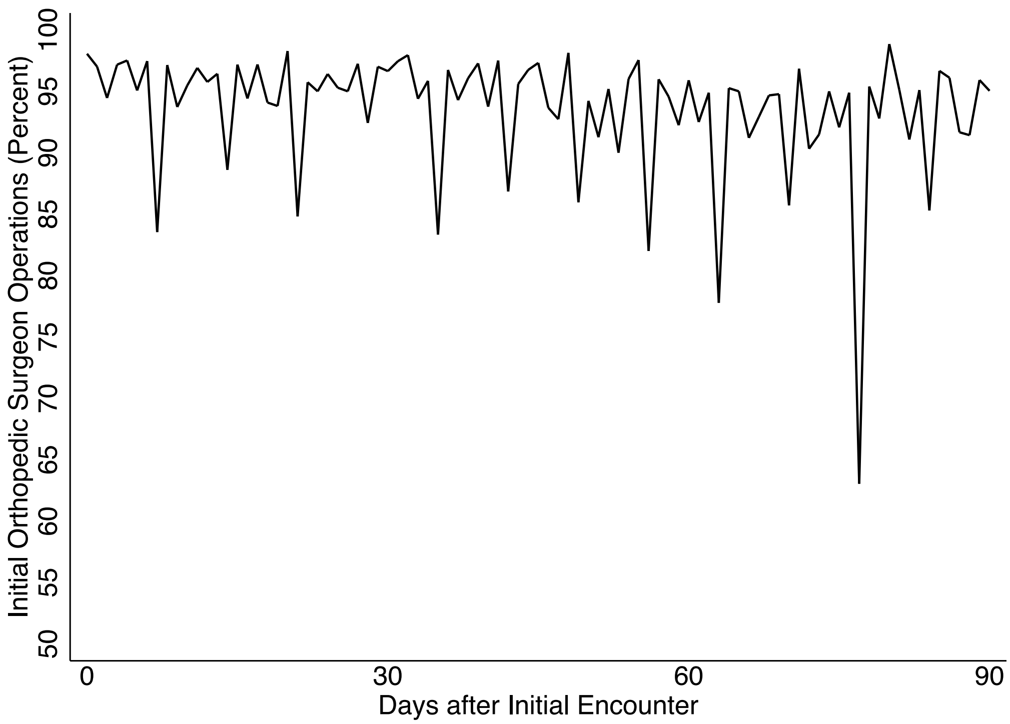
1.3.1 Setting

I study encounters of patients and orthopedic surgeons at a leading academic hospital. These encounters take place at the surgeon's office, at the emergency department, or at another area of the hospital. Office encounters are scheduled with the surgeon's staff for patients who do not require immediate attention. In this chapter, I focus on office encounters constituting the first time that a patient sees any orthopedic surgeon at the hospital. The resulting data comprises 83,595 encounters by 20 attending orthopedic surgeons conducted from 2003–2012.[3]

During an initial office encounter, the orthopedic surgeon becomes acquainted with the patient, determines the diagnosis, and discusses viable treatment options. The surgeon may request X-rays and tests necessary to identify or confirm the diagnosis. In some cases, the physician performs minor procedures — such as an injection — or prescribes medications. At the end of the encounter, the patient may be asked to schedule a follow-up encounter at the surgeon's office or an operation appointment with the physician's staff.

In this chapter, I am primarily interested in the operative decision made by the patient and the orthopedic surgeon after their initial encounter. I track whether the patient is operated on within 30, 90, and 360 days of the initial office appointment. The patient may be operated on by the physician conducting the initial office encounter or by another orthopedic surgeon practicing at the hospital. Figure 1.4 highlights that more than 90 percent of operations are performed by the orthopedic surgeon who conducted the initial office encounter. This ratio does not change substantially over time. Thus, I focus on operations performed by the initial orthopedic surgeon for the remainder of this chapter. Results are robust to including operations performed by all orthopedic surgeons working at at the hospital.

[3]The data excludes orthopedic surgeons maintaining an office outside of the hospital because I cannot track their initial patient encounters. The data also excludes orthopedic surgeon performing fewer than 1,000 initial patient encounters or 100 operations during the observation period.



Notes: This figure shows the percentage of operations performed by orthopedic surgeons conducting the initial office encounter. I also observe whether a patient is operated on by another orthopedic surgeon practicing at the same hospital. The percentage of operations performed by the initial orthopedic surgeon is computed individually for each day following the initial office encounter.

Figure 1.4: *Operations by Initial Orthopedic Surgeon*

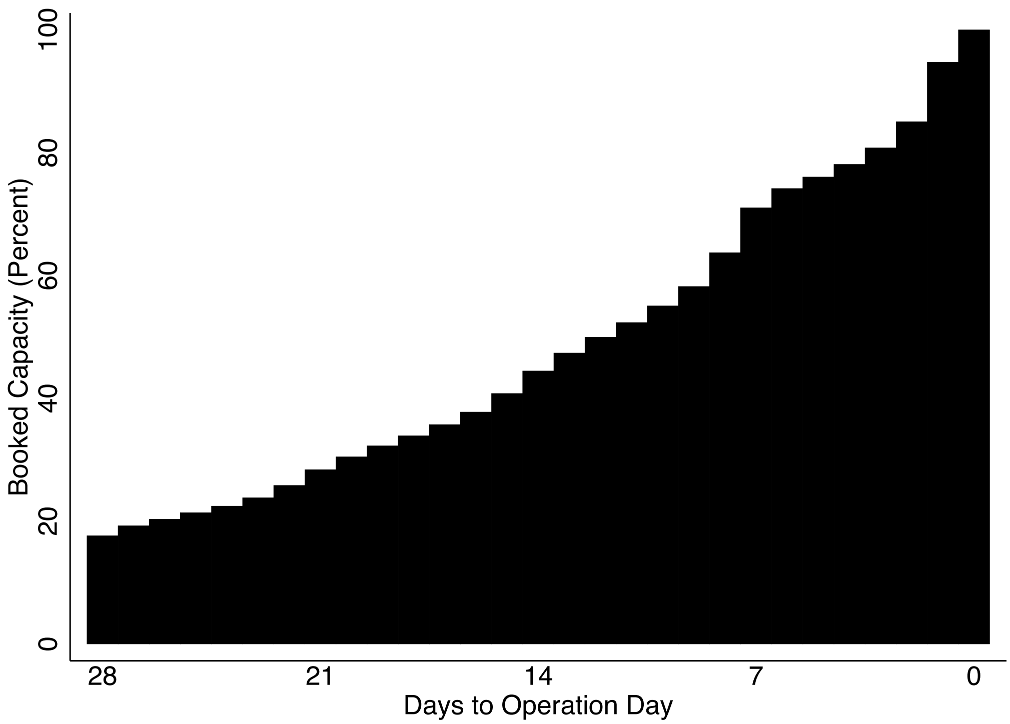
The setting studied in this chapter offers two distinct advantages. First, operative decisions in orthopedic surgery are both numerous and more discretionary than in other fields in medicine. Orthopedic surgeries account for more than 30 percent of all operations conducted at the hospital analyzed here. This percentage — which is the highest among all of the hospital's departments — is fairly typical for a large academic institution. Meanwhile, scientific evidence on the optimal management of many orthopedic conditions is rare, and the benefits of surgical treatment versus non-operative management are idiosyncratic to patients. Consequently, physicians exert substantial discretion over their operative decisions and patients may often be given a choice between surgery and conservative treatment.

Second, a large percentage of orthopedic conditions do not require immediate treatment, affecting the scheduling process of both appointments and operations. As most of their patients do not require urgent attention, orthopedic surgeons typically see new patients during an office appointment scheduled by the surgeon's secretary on a first-come, first-served basis. Thus, physicians have minimal knowledge about their new patients before their initial encounter and do not prioritize based on severity. This feature facilitates my identification strategy discussed in Section 1.4. Operations are similarly added to a surgeon's schedule based on the available capacity. This strategy is feasible because operative delays are frequently acceptable. The next subsection discusses the operation scheduling process and the resulting variation of physician capacity in much more detail.

1.3.2 Exogenous Variation of Physician Capacity

At the hospital considered in this study, orthopedic surgeons have to schedule operation days — but not individual operations — months in advance. Physicians scheduling an operation day are guaranteed to have access to a dedicated operation room and the staff necessary to perform an operation. This staff typically includes residents, anesthesiologists, scrub nurses, and circulating nurses. Patients not requiring emergency surgery — including all the patients considered in this chapter — are typically operated during an operation day. Meanwhile, emergency cases are frequently treated in dedicated emergency operating rooms.

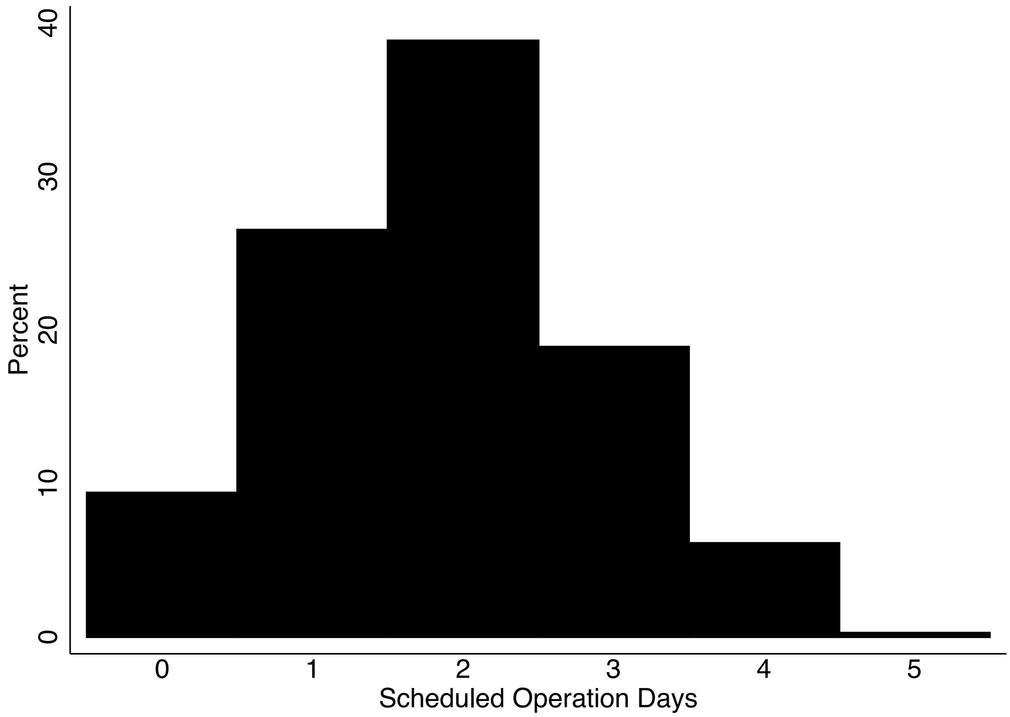
Individual operations are scheduled within a shorter time frame. Physicians typically book operations within weeks or even days. Figure 1.5 is based on the number of operations performed during an operation day and shows the percentage of these operations that have already been booked at a given day before the operation day. It includes the operation days of all relevant orthopedic surgeons during the observation period. On average, more than 80 percent of capacity is still available four weeks before the operation day. This percentage reduces to about 55 percent within two weeks before the operation day. Even within a week before the operation day, more than 20 percent of capacity is still available.



Notes: This figure shows the percentage of operations that have already been booked within a given number of days before the operation day. The data is averaged across the operation days of all orthopedic surgeons.

Figure 1.5: *Booked Capacity*

This study relies on data capturing the time when an operation was scheduled, performed, or cancelled. This information is available for any surgery involving one of the relevant surgeons during the entire observation period. Based on this data it is straightforward to track operations that have been scheduled but not yet executed. By contrast, I have to impute scheduled operation days, implying that I observe an operation day as soon as at least one operation has been scheduled during that day. To mediate this issue, I focus on the expected operation schedule within seven days of an initial office encounter. Within this time frame, orthopedic surgeons rarely keep an operation day entirely open and are keenly aware of their available capacity. Results are robust to using the operation schedule within 14 days of an initial office appointment.

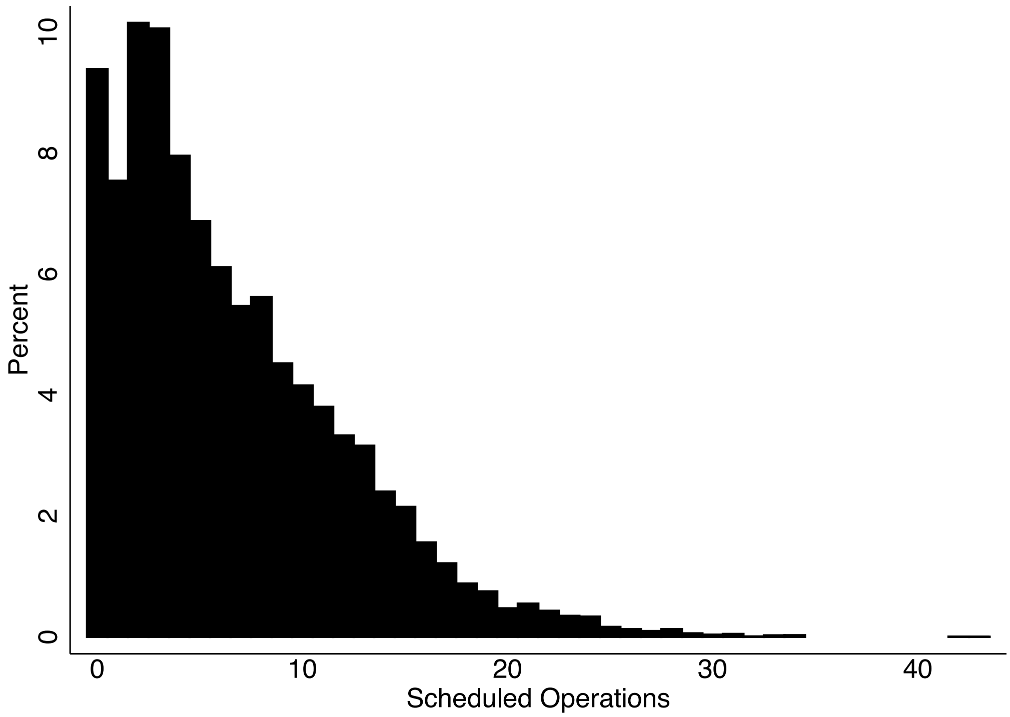


Notes: This figure shows the distribution of the number of scheduled operation days within seven days of an initial office encounter.

Figure 1.6: *Scheduled Operation Days*

Figure 1.6 plots the distribution of scheduled operation days within seven days of the initial office encounters studied in this chapter. On average, the relevant physicians schedule 1.87 operation days within seven days of an initial office encounter considered in this chapter. The number of scheduled operation days varies substantially, ranging from from zero to five in about ten and one percent of relevant encounters, respectively. Presumably, this variation stems from physicians' vacations, conference attendances, and other external obligations scheduled well in advance.

The distribution of scheduled operations within seven days of a relevant encounter is shown in Figure 1.7. The average number of scheduled operations within seven days of a relevant encounters is 6.58. The median duration of an operation that is actually performed is 58 minutes. Physicians have already scheduled 14 or more operations within the next



Notes: This figure shows the distribution of the number of scheduled operations within seven days of an initial office encounter.

Figure 1.7: *Scheduled Operations*

week during ten percent of relevant encounters. Meanwhile, another ten percent of relevant encounters are not linked to any upcoming operation during the next week. Obviously, the number of scheduled operation days has a major impact on the number of scheduled operations. The correlation of these two values amounts to 0.64. In addition, the number of scheduled operations varies with the number and severity of recently seen patients. The correlation between the number of scheduled operations and the number of new patients seen by same physician within the last 28 days is 0.14.[4]

[4]The number of new patients seen by the same physician within the past 28 days is constructed based on the sample of 83,595 encounters analyzed in this chapter.

1.3.3 Measuring Physician Capacity

Ideally, a measure of physician capacity should capture the number of open surgery slots across the relevant operation days. Unfortunately, the number of surgery slots during a given operation day is only observed retrospectively. To address this issue, I combine the number of scheduled operations and scheduled operation days to estimate the number of open surgery slots. I impute physician capacity — the total number of surgery slots — based on the average number of operations conducted during an operation day.

The average capacity per operation day is computed individually for each physician across the entire observation period. Formally this can be described as follows:

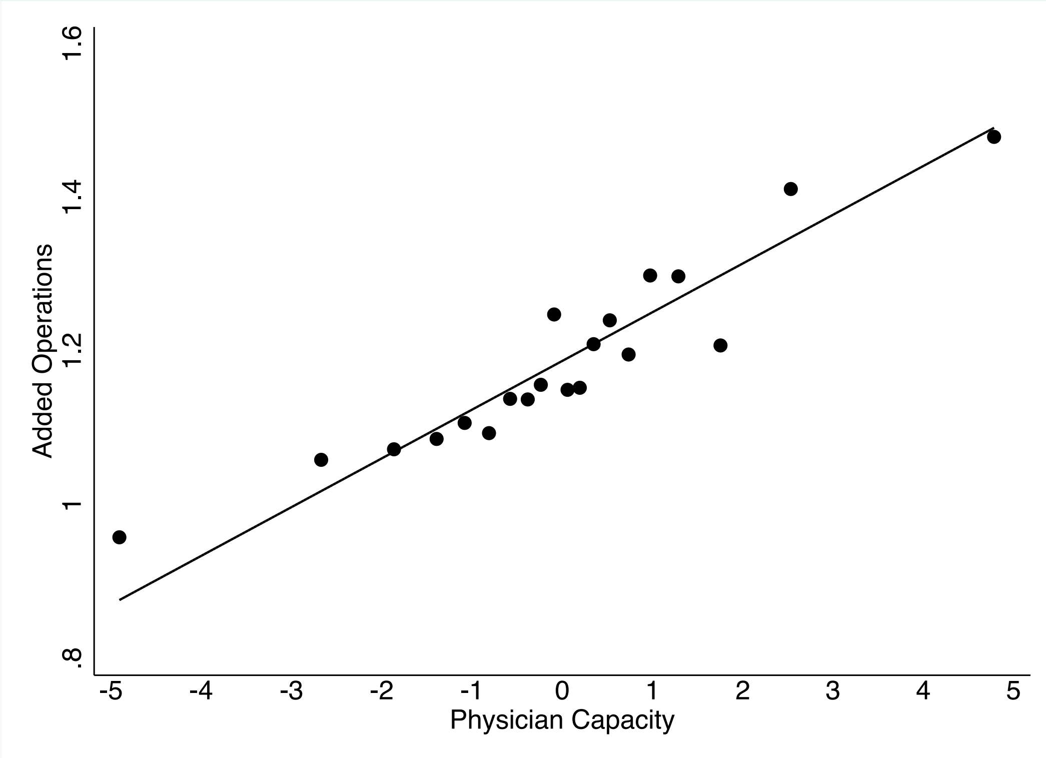
$$\text{Average Capacity}_p = \frac{\text{Operations}_p}{\text{Operation Days}_p} \tag{1.1}$$

Based on this measure, I determine the total capacity of physician p within time period t of a relevant encounter. This requires multiplying the average capacity of physician p with the number of scheduled operation days during time period t. Subsequently, I subtract the number of operations that physician p already scheduled during time period t. The resulting number captures how many operations can still be booked by physician p within time period t, assuming the orthopedic surgeons operates at average capacity. This yields:

$$C_{pt} = \text{Average Capacity}_p * \text{Scheduled Operation Days}_{pt} - \text{Scheduled Operations}_{pt} \tag{1.2}$$

Here, C denotes the available capacity of physician p during time period t, the crucial explanatory variable of this chapter. As noted previously, t is set to seven days to mediate concerns about operation days that have been scheduled but cannot be observed in the data used for this study. Meanwhile, C is defined at the beginning of the day of the focal initial office appointment. This definition addresses the potential endogeneity of physician capacity, ensuring that C is not affected by the operative decision made by the focal patient.

To check whether C reflects the physician capacity that is actually realized — and can be observed retrospectively — I analyze how well capacity predicts the number of operations added to a physician's schedule within seven days after a relevant encounter. Figure 1.8



Notes: This figure shows the relationship between physician capacity at the time of an initial office encounter and the number of operations added to the surgeon's schedule within seven days after the corresponding encounter. The figure adjusts for physician-month fixed effects.

Figure 1.8: *Physician Capacity and Added Operations*

plots the relationship of C and added operations across 20 bins, adjusting for physician-month fixed effects.[5] The relationship is almost perfectly linear, suggesting that C predicts actual physician capacity with considerable precision. At the same time Figure 1.8 shows that the C does not exactly reflect available capacity. Within seven days, C is frequently negative, suggesting that orthopedic surgeons are more than fully booked based on their average capacity. Nevertheless, these physicians seem to still add some operations to their schedule. On the other hand, many orthopedic surgeons seem to add far fewer operations to their schedule than C would permit. An additional unit of C does not translate to

[5] These estimates are based on a OLS regression of added operations on fixed effects for the each of the 20 quantiles of physician capacity and each physician-month.

analogous increase in the number of operations actually added to the physician's schedule. The obvious explanation is that my measure of physician capacity does not account for a surgery's expected duration, but averages across all of a surgeon's operations without taking into account the type of surgery.

1.3.4 Data

I study 83,595 initial office encounters conducted by 20 orthopedic surgeons from 2003–2012. The data captures any office encounter between a relevant orthopedic surgeon and a patient that has not previously been seen by the same or another orthopedic surgeon at the same hospital. For each encounter, I observe the time when the encounter took place. I also have access to the patient's age, gender, and race. Diagnoses made during the encounter are captured by ICD-9 codes. In addition, the data tracks whether the patient was insured by a fee for service provider (FFS), a managed care organization (MCO), Medicare, or Medicaid.

To measure the patient's socioeconomic status, I use the patient's ZIP code to supplement the data with ZIP code characteristics. I include the percentage of the ZIP code population which is white, lives in a rural area, is aged 65 and older, or is aged 25 and older and has at least a high school degree. I also include the median household income at the ZIP code level. For each ZIP code characteristic, missing values are replaced with the sample mean and indicated in a separate variable. I obtain these measures from the 2012 American Community Survey and the 2010 Census provided by the US Census Bureau.

Table 1.1 provides summary statistics across quintiles of physician capacity computed individually for each physician. Physician capacity fluctuates substantially, and operation rates steadily increase with available capacity. Across all time periods, the operation rate in the top quintile of physician capacity is 1.5 percentage points higher than the operation rate in the bottom quintile. This effect is statistically significant at the five percent level. The increase of the operation rate is especially pronounced between the fourth and fifth quintile, suggesting that the raw relationship between physician capacity and operation rates is nonlinear. Physician characteristics and the timing of encounters are fairly balanced.

Table 1.1: *Summary Statistics by Physician Capacity Quintile*

	Physician Capacity Quintile				
	1	2	3	4	5
Physician Capacity					
7 Days	-3.004	-1.078*	-0.016*	1.026*	2.862*
Operation Rates					
30 Days	0.081	0.082	0.085	0.087	0.096*
90 Days	0.164	0.166	0.167	0.171	0.179*
360 Days	0.221	0.223	0.227	0.228	0.236*
Patient Characteristics					
Age	38.64	38.66	38.51	38.52	38.62
Percent Male	0.510	0.507	0.515	0.519	0.512
Percent White	0.842	0.846	0.839	0.841	0.837
Insurance Types					
Percent FFS	0.161	0.166	0.168	0.167	0.168
Percent MCO	0.634	0.633	0.633	0.634	0.630
Percent Medicaid	0.031	0.030	0.029	0.030	0.031
Percent Medicare	0.162	0.158	0.156	0.156	0.160
Percent Uninsured	0.013	0.013	0.014	0.013	0.011
ZIP Code					
Median Household Income	79659	79657	79530	79504	79759
Percent Age $\geq$ 65	0.141	0.141	0.141	0.141	0.141
Percent Highschool	0.905	0.905	0.904	0.905	0.905
Percent White	0.818	0.818	0.817	0.818	0.819
Percent Rural	0.080	0.077	0.079	0.078	0.078
Timing					
Day Of Week	2.717	2.750	2.747	2.757	2.728
Day Of Month	15.58	15.72	15.35	15.65	15.58
Hour Of Day	10.93	10.90	10.88	10.86	10.92
Observations	16719	16719	16719	16719	16719

*Notes: ZIP code information is reported for observations without missing ZIP code information. * denotes significant difference from the bottom quintile at the five percent level, computed using standard errors clustered at the physician level.*

1.4 Empirical Strategy

1.4.1 Identification

I explore whether physician capacity affects operative decisions. If an operative decision is made in the patient's best interest, physician capacity should not systematically vary with operation rates unless it is correlated with unobservable patient and physician characteristics. I rely on two strategies to ensure that this identifying assumption is met. First, I focus on the initial encounter between a patient and a physician. In this setting, the physician has very limited unobservable information about the patient. Meanwhile, the patient does not observe the physician's schedule. Consequently, I argue that patient characteristics are unlikely to systematically fluctuate with available physician capacity. Second, I rely on variation of the operative decisions made by the same physician within short periods of time. Physician expertise should not vary within these time periods.

Both physicians and patients might consider unobservable patient characteristics when scheduling an office appointment, thus introducing a relationship between unobservable patient characteristics and physician capacity. I address this issue by focusing on the initial office appointment of a patient that has never been seen by any orthopedic surgeon practicing at the same hospital. By definition, the physician has never seen the patient when scheduling the appointment. In fact, the initial office appointment is typically scheduled with the physician's secretary, implying that physician is not even aware of the patient at the time when the appointment is scheduled. As the patients considered in this chapter schedule an office appointment rather than reporting to the emergency room, they do not require immediate attention, suggesting that the physician would not prioritize their appointment based on patient characteristics. All in all, it appears fairly unlikely that physicians plan their office appointment schedule based on patient characteristics.

When scheduling an office appointment, patients decide between several time slots offered by the physician's secretary. I hypothesize that unobservable patient characteristics influence timing preferences. For instance, a patient who is working presumably prefers an appointment in the afternoon or on Friday. To address this bias, I include fixed effects for the hour of the day, the day of the week, and the day of the month of the time when the appointment is performed. Another source of bias is the urgency patients exhibit in obtaining an appointment. If the patient wishes to be seen as soon as possible, he is likely to accept the earliest appointment available. It is unclear whether the physician has less capacity available during these appointment. If this was the case, and if patients exhibiting more urgency were more likely to require surgery, this would bias the relationship between physician capacity and operation rates downwards. Thus, a positive relationship between physician capacity and operation rates could not be rationalized by differences in patients' urgency to obtain an appointment.

Perfectly adjusting for physician expertise is an insurmountable challenge. Several approaches have been developed to circumvent this issue. A large literature compares operation rates across regions rather than across physicians (Glover 1938, Wennberg and Gittelsohn 1973, Fisher et al. 2003), implicitly assuming that physician characteristics balance out across areas. However, recent research casts doubt on the validity of this approach (Chandra and Staiger 2007). In the absence of exogenous variation of patient characteristics — such as in Doyle et al. (2010) — there are no reliable methods to disentangle the impact of patient and physician characteristics, respectively, on differences in treatment decisions and health outcomes.

I address this issue by estimating the relationship between physician capacity and operation rates based on variation of operative decisions made by the same physician during the same month using physician-month fixed effects. Given that all of the orthopedic surgeons considered in this chapter have multiple years of experience, another month of experience should not have a significant impact on physician expertise. This approach follows a small but growing literature exploiting variation of treatment choices within rather than across physicians. For instance, Choudhry et al. (2006) explore whether adverse events affect the same physician's subsequent medication behavior.

1.4.2 Estimation

The estimating equation underlying my analyses can be formalized as follows:

$$O_{ij} = \alpha + \beta C_{j(i)} + \gamma \mathbf{X}_i + \delta \mathbf{Z}_{z(i)} + \eta \mathbf{P}_{j(i)} + \theta \mathbf{T}_{t(i)} + \epsilon_i \tag{1.3}$$

O indicates whether patient i is operated by physician j within a given time period after their initial office appointment. In my analysis, I compute whether the patient was operated within 30, 90, and 360 days of the initial office appointment, respectively. C reflects the capacity physician j has available to operate patient i within one week after first seeing patient i.

I control for a wide range of covariates. $\mathbf{X}$ accounts for patient characteristics comprising indicators for five-year age groups, gender, race, ICD-9 diagnoses codes, and insurance types. $\mathbf{Z}$ denotes a vector of the characteristics of ZIP code z patient i is living in. $\mathbf{P}$ controls for physician-month fixed effects, and $\mathbf{T}$ includes indicators for the day of the week, day of the month, and hour of the day of time t when the initial office appointment is performed. I estimate Equation 1.3 using a linear probability model to simplify the interpretation of coefficients. The results are robust to using a probit model. Standard errors are clustered at the physician-month level.

1.5 Results

1.5.1 Physician Capacity

Table 1.2 presents estimates of the impact of physician capacity on operation rates within 30, 90, and 360 days of a patient's initial office appointment. I find that physician capacity has a positive and statistically significant impact on operation rates. My finding suggests that each unit of physician capacity — reflecting an additional open surgery slot within seven days of the initial encounter — increases a patient's probability of being operated on between 0.1 and 0.2 percentage points. These result are largely unaffected by controlling for a large set of patient and encounter characteristics.

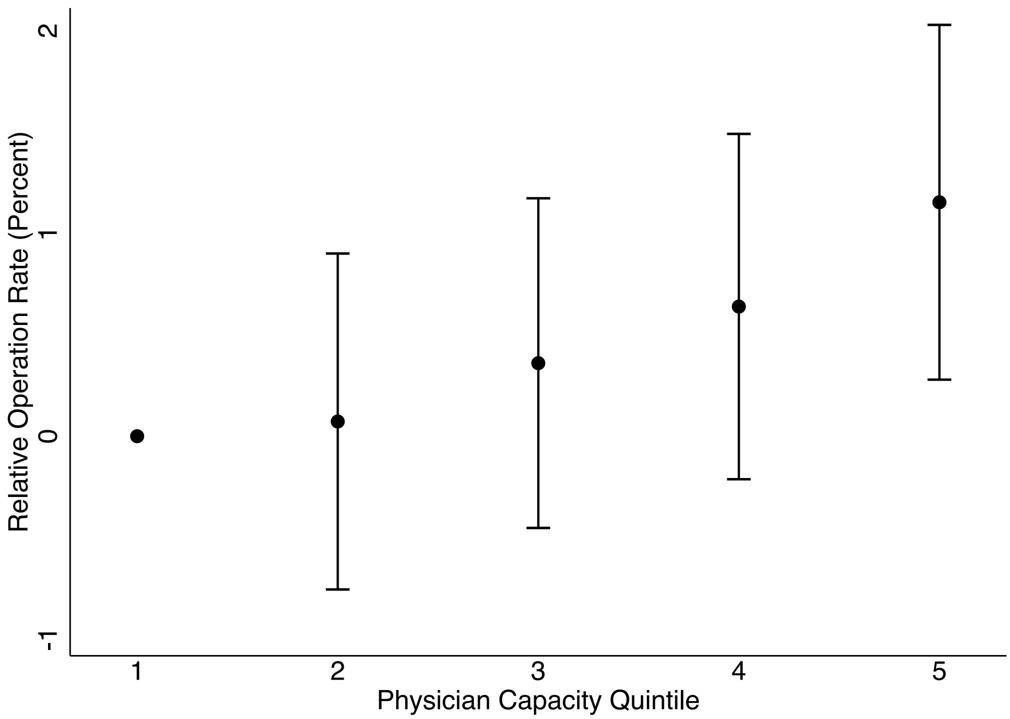
Table 1.2: *Physician Capacity and Operative Decisions*

	Operation					
	30 Days		90 Days		360 Days	
	(1)	(2)	(3)	(4)	(5)	(6)
Physician Capacity	0.163***	0.126**	0.238***	0.215***	0.171**	0.153**
	(0.057)	(0.052)	(0.069)	(0.065)	(0.076)	(0.069)
Physician Month	Yes	Yes	Yes	Yes	Yes	Yes
Patient Characteristics	No	Yes	No	Yes	No	Yes
ICD-9 Codes	No	Yes	No	Yes	No	Yes
Insurance Types	No	Yes	No	Yes	No	Yes
ZIP Code	No	Yes	No	Yes	No	Yes
Timing	No	Yes	No	Yes	No	Yes
Observations	83595	83595	83595	83595	83595	83595
Adjusted R2	0.034	0.161	0.057	0.168	0.079	0.187
Mean Operation Rate	8.596	8.596	16.928	16.928	22.701	22.701

*Notes: This table reports the effect of physician capacity on whether a patient was operated within 30, 90, or 360 days of an initial office encounter. Coefficients are multiplied by 100 and can be interpreted as percentages. Models are estimated by OLS and standard errors are clustered at the physician-month level. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

Increasing physician capacity by one standard deviation — or 2.5 units of physician capacity — is associated with an operation rate that is between 0.3 and 0.6 percentage points higher according to the estimates reported in Table 1.2. Relative to the mean operation rate, a patient's likelihood of being operated on increases by 3.7 percent within a month, 3.2 percent within a quarter, and 1.7 percent within a year. Given that I consider the entire patient population seen by orthopedic surgeons, these increases are substantial.

To test whether the relationship between physician capacity and operation rates is nonlinear, I use indicators for the quintile of physician capacity rather than the level of physician capacity to compute (1.3). The resulting estimates are plotted in Figure 1.9. Operation rates are normalized so that the operation rate of the bottom quintile of physician capacity is zero. Brackets mark 95 percent confidence intervals. Relative to the lowest quintile of physician capacity, operation rates steadily increase with available physician capacity. Compared to the lowest quintile, operations rates in the top quintile are 1.2 percentage



Notes: This figure shows the operation rate within 90 days of an initial office encounter relative to the bottom quintile of physician capacity. Coefficients are plotted as circles. Vertical brackets show 95 percent confidence intervals. The underlying model is estimated by OLS and controls for physician-month, patient characteristics, ICD-9 codes, insurance types, ZIP code, and timing fixed effects. Standard errors are clustered at the physician-month level.

Figure 1.9: *Operation Rate by Physician Capacity Quintile*

points higher, translating to a 7.1 percent increase over the mean 90 day operation rate. Operations rates are significantly different from the bottom quintile only in the top quintile.

1.5.2 Physician Discretion

Presumably, the discretion physicians exert when deciding whether to operate on a patient varies across conditions. I hypothesize that physician capacity affects operative decisions more when orthopedic surgeons have substantial discretion. Unfortunately, consistently measuring physician discretion across all diagnoses is almost impossible. To address this issue, I focus on seven of the most frequent diagnoses in my data. Across these diagnoses, both the medical literature and discussions with physicians enable me to assess the extent of physician discretion. For each of these diagnoses, I estimate a simple version of Equation 1.3 adjusting for physician-fixed effects. Figure 1.10 plots the resulting estimates. The constant term — describing the operation rate when physician capacity is zero — is plotted as circle. The coefficient for capacity is multiplied by the standard deviation of physician capacity, and the resulting range of operation rates is represented as horizontal bracket.

Figure 1.10 demonstrates that the impact of physician capacity on operation rates varies across diagnoses and appears to fluctuate systematically with physician discretion. Patients with a carpal tunnel syndrome[6], hand osteoarthrosis[7], or a trigger finger[8] are most frequently seen by hand surgeons. While there are relatively clear guidelines when to operate on a patient with a carpal tunnel syndrome, indications to operate a patient with a trigger finger are somewhat less clear. By contrast, orthopedic hand surgeons have substantial discretion in whether to operate a patient with hand osteoarthrosis. Figure 1.10 reflects these facts, indicating that physician capacity has almost no influence on carpal tunnel operation rates, but matters enormously for hand osteoarthrosis operation rates.

Other frequent upper extremity diagnoses include rotator cuff tears[9] — the most common shoulder condition treated by orthopedic upper extremity surgeons — and Colles fractures[10] involving the arm. The medical literature suggests that there is substantial discretion whether to operate on patients suffering from a rotator cuff tear (Williams et al. 2004). Figure 1.10 again mirrors this fact, indicating that physician capacity impacts rotator cuff

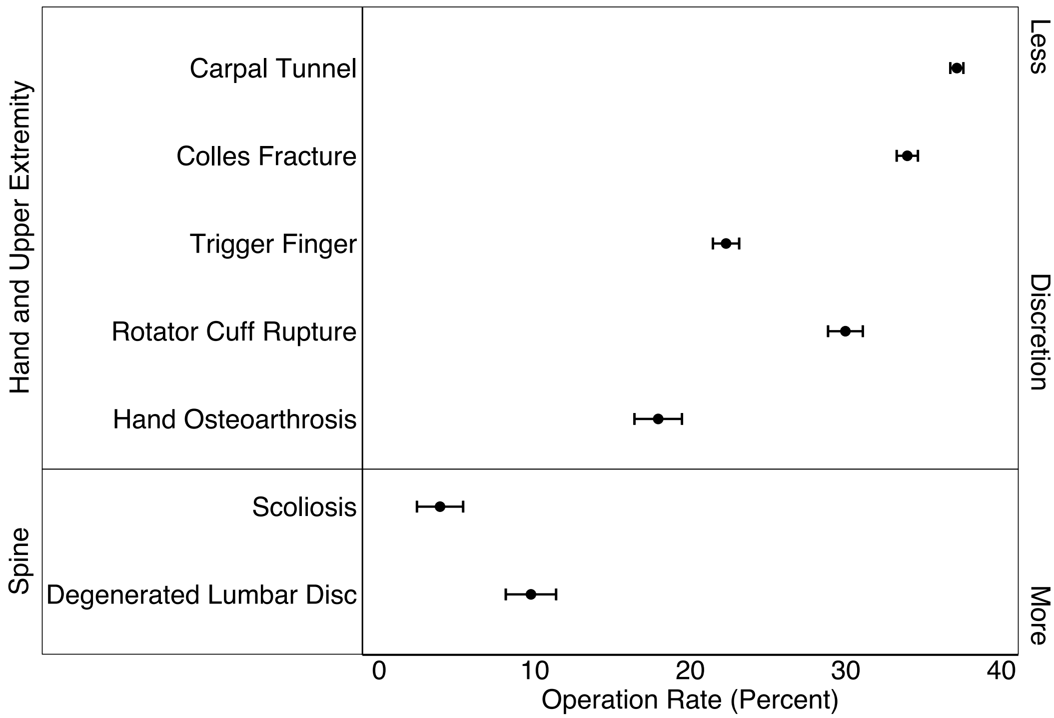
[6] A carpal tunnel syndrome occurs if the median nerve — running from the axilla to the hand — becomes pressured. This condition can be treated by surgically releasing the ligament pressuring the median nerve.

[7] Hand osteoarthritis refers to joint degradation in the hand, causing pain and stiffness. It is treated non-surgically by injecting or splinting and surgically by a number of procedures such as proximal row carpectomy and total wrist arthroplasty or arthrodesis.

[8] A patient has a trigger finger if the corresponding finger flexor tendon is locked, snapped, or catched. To treat the resulting pain and dysfunction non-operatively, physicians may either splint or inject the finger. The surgical treatment option requires cutting the sheath restricting the tendon.

[9] Rotator cuff tears involve tearing a tendon of the rotator cuff muscles. Tears may be traumatic or chronic. Both operative and conservative treatments — involving physiotherapy and pain management — are feasible.

[10] A fracture of the distal radius, the Colles fracture typically affects patients with osteoporosis. Displaced Colles fractures are treated surgically, while non displaced Colles fractures may be managed conservatively.



Notes: This figure shows the operation rate within 90 days of an initial office encounter across frequent diagnoses. Circles indicate the estimated operation rate when physician capacity is zero. Horizontal brackets indicate the impact of a one standard deviation move in physician capacity above or below zero. The underlying models estimate the impact of capacity on operation rates across all patient with a given diagnosis using OLS and adjust for physician-month fixed effects.

Figure 1.10: *Operation Rate across Frequent Diagnoses*

tear operation rates. Meanwhile, the Colles fracture — which allows for less physician discretion — is less affected by physician capacity.

Spine surgery is often seen as the most discretionary discipline within orthopedic surgery. Not surprisingly, the operation rates for scoliosis[11] and degenerated lumbar disc[12] — the most frequent spine-related diagnoses in my data — are affected very substantially by

[11]Scoliosis is a condition where the patient's spine is curved abnormally, bending from side to side. Frequently, operative treatment is not warranted. Other treatment options include physiotherapy, casting, or bracing. Surgical options include anterior and posterior fusion.

[12]Degenerated lumbar discs in the spine can lead to enormous chronic pain for some patients. They may be treated conservatively with physiotherapy, medications reducing inflammation, and spinal injections. Potential operations include discectomy and fusion, laminectomy, and corpectomy.

physician capacity. The ratio between the lowest and highest operation rate for scoliosis is more than two, suggesting that the operation probability more than doubles following a two standard deviation increase in physician capacity.

1.5.3 Physician Remuneration

A physician's available capacity may not only influence the surgeon's operative decisions, but also impact the reimbursement physicians receive for treating a patient. While the data used for this chapter does not comprise actual payments made to physicians, it tracks all procedures administered by a relevant orthopedic surgeon. I use the corresponding CPT codes to compute the relative value units (RVU) associated with all of the treatments provided by the focal physician performing the initial appointment with a patient within 30, 90, and 360 days of the initial encounter. RVUs are used by Medicare to compute the reimbursement physicians receive for a given procedure, but should not be interpreted in dollar terms.[13]

I estimate the impact of physician capacity on RVUs analogous to Equation 1.3. Table 1.3 presents the corresponding results. Columns 1, 3, and 5 show the impact of physician capacity on RVUs. Increasing physician capacity by one standard deviation raises RVUs by 1.7, 2.1, and 1.2 percent within 30, 90, and 360 days of the initial encounter, respectively. It is noteworthy that the estimates across 30 and 360 days are fairly noisy.[14]

I am primarily concerned with the impact of operative decisions on the relationship between physician capacity and RVUs. Columns 2, 4, and 6 demonstrate that operative decisions entirely explain the impact of physician capacity on RVUs. The point estimates

[13]RVUs have to be multiplied by the geographic practice cost index varying across regions and the Medicare conversion factor — which is constant across regions — to compute the reimbursement in dollar terms.

[14]A potential explanation for this finding is that not all procedures are captured by CPT codes. A considerable percentage of operations is captured solely by ICD-9 procedure codes. Both CPT and ICD-9 procedure codes may be used for billing purposes. As there is no generally accepted translation from ICD-9 procedure codes to CPT codes, I refrain from using these codes. Consequently, RVUs are biased downwards, especially at the time when an operation captured by ICD-9 procedure codes has already been conducted, but the corresponding after-care — measured by CPT codes — has yet to be performed. This phenomenon might help to explain why the coefficient in Column 1 of Table 1.3 is so low.

Table 1.3: *Physician Capacity and Relative Value Units*

	Relative Value Units					
	30 Days		90 Days		360 Days	
	(1)	(2)	(3)	(4)	(5)	(6)
Physician Capacity	0.035	-0.005	0.089***	0.002	0.073*	-0.004
	(0.022)	(0.018)	(0.031)	(0.025)	(0.041)	(0.035)
Operation 30 Days		31.826***				
		(0.499)				
Operation 90 Days				40.523***		
				(0.596)		
Operation 360 Days						50.497***
						(0.774)
Observations	83595	83595	83595	83595	83595	83595
Adjusted R2	0.137	0.422	0.176	0.469	0.208	0.479
Mean RVU	5.033	5.033	10.308	10.308	15.784	15.784

*Notes: This table reports the effect of physician capacity on the RVUs within 30, 90, or 360 days of an initial office encounter. Models are estimated by OLS and control for physician-month, patient characteristics, ICD-9 codes, insurance types, ZIP code, and timing fixed effects. Standard errors are clustered at the physician-month level. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

for the impact of physician capacity are very close to zero across all time periods. Given that physician capacity is defined based on the number of operations a physician could conduct when operating at average capacity, these findings are expected. In addition, the result that physician capacity has no impact on all but operative decisions strengthens the argument that there is no relationship between physician capacity and unobservable patient characteristics, supporting my identification strategy.

1.5.4 Robustness

I consider three types of robustness checks. I start by estimating Equation 1.3 separately for different groups of patients and for different groups of encounters based on the time of the day and day of the week when the focal encounter took place. In addition, I use each of the sources of variation underlying physician capacity — scheduled operations and scheduled operation days — to confirm that they individually and jointly predict operation rates.

Patient Characteristics Table 1.4 reports the impact of physician capacity on 90 day operation rates across various patient characteristics. As shown in Columns 1–5, physician capacity appears to affect operative decisions differently across age groups. While physician capacity has no discernible effect for patients aged between 20–39 years and very old patients, it significantly impacts the operation rates of patients who are older than 40 years and younger than 80 years. Anecdotal evidence supports this finding, suggesting that operative decisions involving substantial physician discretion frequently affect patients in that age group, especially if they are older than 60 years but still young enough to cope with the risks posed by an operation.

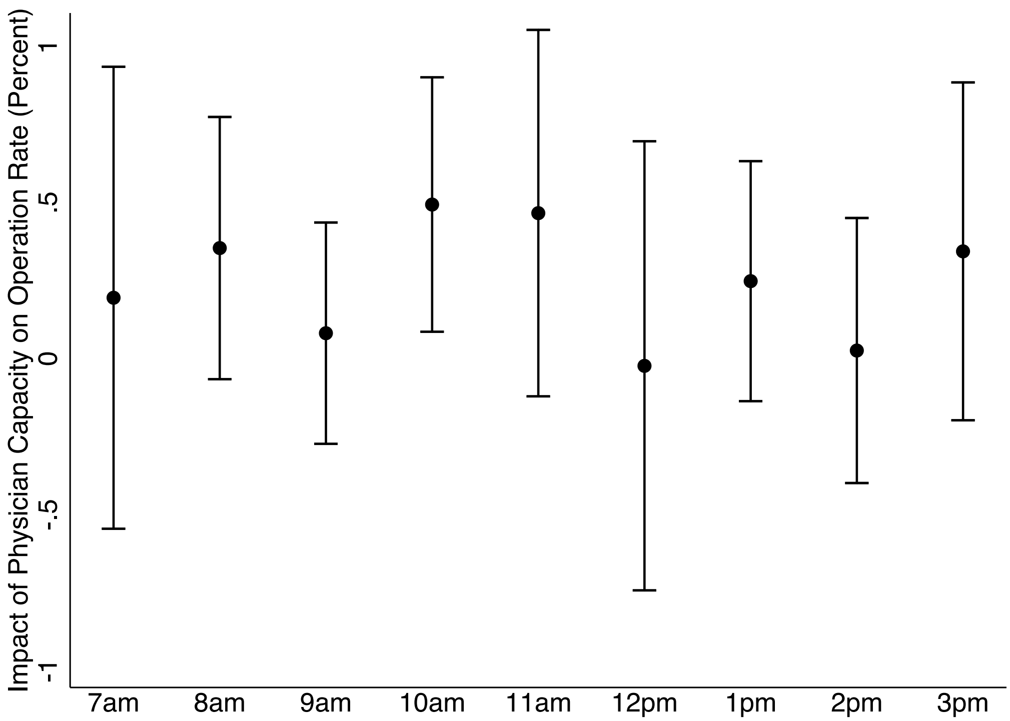
Columns 6–10 present estimates across race. White patients constitute 84 percent of all patients considered in this study. Thus, it is not surprising that the coefficient in Column 10 corresponds to the overall estimates in Table 1.2. While operation rates seem to differ substantially across race, I lack power to identify a differential impact of physician capacity on operative decisions. Similarly, estimates across insurance type presented in Columns 11–15 suffer from insufficient sample size to allow for reliable inferences. Nevertheless, Column 13 at least suggests that patients insured by Medicaid are much more affected by changes in physician capacity than any other group. Meanwhile, patients insured by a MCO are impacted least by physician capacity. This finding is consistent with anecdotal evidence pointing to a more limited role of the discretion exerted by physicians when they are consulting and treating patients insured by a MCO.

Finally, Columns 16–20 look at the impact of the median household income of the patient's ZIP code of residence on the relationship between physician capacity and operation rates. Interestingly, patients living in the lowest quintile by median ZIP household income appear to be twice as affected by physician capacity as the overall estimates in Table 1.2 would imply. Potentially, physicians hesitate to operate on these patients unless they have ample spare capacity because they impute their potential reimbursement from the patient's socioeconomic status.

Table 1.4: *Physician Capacity and Operation Rates across Patient Characteristics*

Dependent Variable: Operation 90 Days					
	Age				
	0–19	20–39	40–59	60–79	80–100
	(1)	(2)	(3)	(4)	(5)
Physician Capacity	0.443*	0.059	0.250**	0.435***	-0.039
	(0.227)	(0.123)	(0.110)	(0.167)	(1.135)
Observations	20501	20608	27078	13606	1798
Mean Operation Rate	8.190	18.711	20.219	20.726	17.853
	Race				
	Asian	Black	Hispanic	Other	White
	(6)	(7)	(8)	(9)	(10)
Physician Capacity	0.103	0.129	0.365	1.235	0.221***
	(0.529)	(0.507)	(0.351)	(1.839)	(0.071)
Observations	3066	3340	5640	1225	70324
Mean Operation Rate	12.883	13.293	12.855	14.694	17.643
	Insurance Type				
	FFS	MCO	Medicaid	Medicare	None
	(11)	(12)	(13)	(14)	(15)
Physician Capacity	0.290	0.118	0.823	0.212	-0.474
	(0.183)	(0.085)	(1.001)	(0.167)	(1.748)
Observations	13887	52895	2518	13247	1048
Mean Operation Rate	17.858	16.225	11.398	19.687	18.511
	ZIP Code Income Quintile				
	1	2	3	4	5
	(16)	(17)	(18)	(19)	(20)
Physician Capacity	0.393**	0.003	0.180	0.237	0.202
	(0.155)	(0.160)	(0.163)	(0.167)	(0.153)
Observations	16334	16534	16393	16248	16811
Mean Operation Rate	16.861	18.368	17.507	16.661	15.169

*Notes: Each cell reports a separate regression. Coefficients are multiplied by 100 and can be interpreted as percentages. Models are estimated by OLS and control for physician-month, patient characteristics, ICD-9 codes, insurance types, ZIP code, and timing fixed effects. Standard errors are clustered at the physician-month level. Observations with missing ZIP code information are dropped from the last panel. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

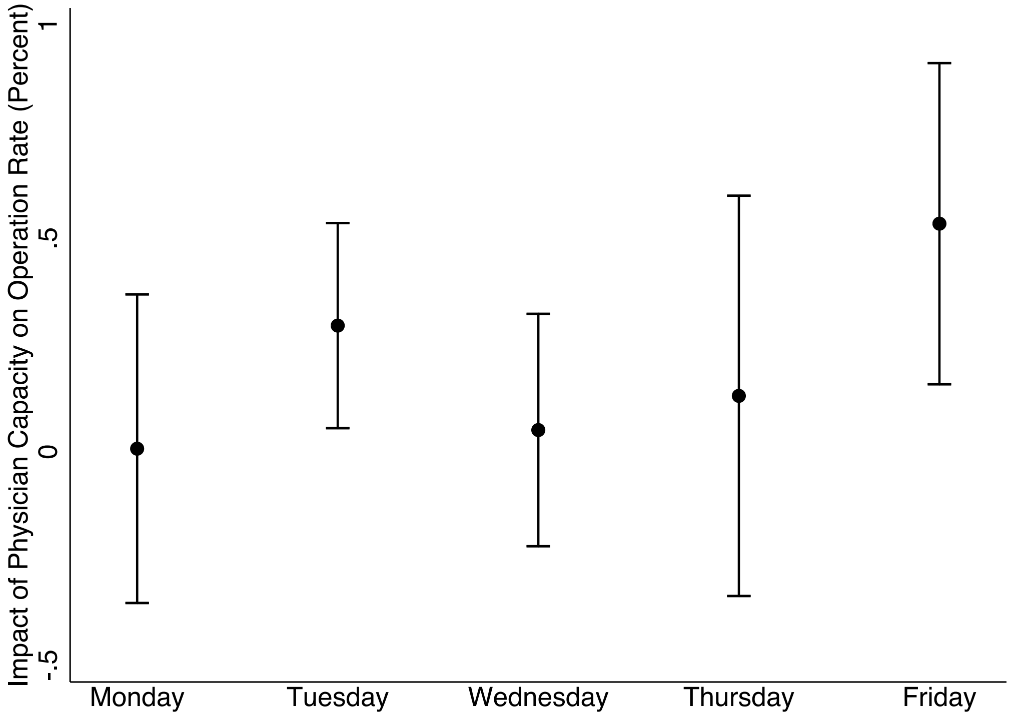


Notes: This figure shows the impact of physician capacity on the operation rate within 90 days of an initial office encounter. Each circle represents the coefficient estimate from a separate regression including all encounters that were performed during that hour. Vertical brackets show 95 percent confidence intervals. The underlying model is estimated by OLS and controls for physician-month, patient characteristics, ICD-9 codes, insurance types, ZIP code, and timing fixed effects. Standard errors are clustered at the physician-month level.

Figure 1.11: *Physician Capacity and Operation Rates across Hours*

Encounter Timing In Figure 1.11, I present estimates of the impact of physician capacity on operative decisions within 90 days of the initial encounter across the hour of the day when the focal encounter took place. Standard errors — shown as vertical brackets — are substantial given the limited amount of data available during each hour. I cannot reject the hypothesis that physician capacity has the same impact on operation rates across all hours.

Nevertheless, it is noteworthy that physician capacity seems to matter most during the late morning and late afternoon. As many orthopedic surgeons schedule office encounters for either the morning or the afternoon, this finding suggests that physicians are especially



Notes: This figure shows the impact of physician capacity on the operation rate within 90 days of an initial office encounter. Each circle represents the coefficient estimate from a separate regression including all encounters that were performed during that weekday. Vertical brackets show 95 percent confidence intervals. The underlying model is estimated by OLS and controls for physician-month, patient characteristics, ICD-9 codes, insurance types, ZIP code, and timing fixed effects. Standard errors are clustered at the physician-month level.

Figure 1.12: *Physician Capacity and Operation Rates across Weekdays*

considering their capacity towards the end of a shift. Figure 1.11 might even underreport this effect as physician capacity is measured at the start of each day. Consequently, actual physician capacity at the end of a shift might be lower than observed physician capacity, implying that the effect of physician capacity on operation rates would be biased downwards.

Figure 1.12 plots the impact of physician capacity on operation rates across weekdays. Again, the statistical power is not sufficient to reject the hypothesis the influence of physician capacity does not vary across weekdays. Taken at face value, the results suggest that orthopedic surgeons consider their capacity twice as much during Fridays. Explanations

Table 1.5: *Sources of Physician Capacity*

	Operation 90 Days					
	(1)	(2)	(3)	(4)	(5)	(6)
Scheduled Operation Days	0.269		0.660***	0.232		0.599***
	(0.174)		(0.243)	(0.167)		(0.232)
Scheduled Operations		-0.021	-0.131**		-0.024	-0.123**
		(0.045)	(0.063)		(0.042)	(0.058)
Physician Month	Yes	Yes	Yes	Yes	Yes	Yes
Patient Characteristics	No	No	No	Yes	Yes	Yes
ICD-9 Codes	No	No	No	Yes	Yes	Yes
Insurance Type	No	No	No	Yes	Yes	Yes
ZIP Code	No	No	No	Yes	Yes	Yes
Timing	No	No	No	Yes	Yes	Yes
Observations	83595	83595	83595	83595	83595	83595
Adjusted R2	0.057	0.057	0.057	0.168	0.168	0.168
Mean Operation Rate	16.928	16.928	16.928	16.928	16.928	16.928

*Notes: This table reports the effect of scheduled operation days and operations within seven days of an initial office encounter on operation rates within 90 days of the focal encounter. Coefficients are multiplied by 100 and can be interpreted as percentages. Models are estimated by OLS and standard errors are clustered at the physician-month level. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

for this finding could range from differences in patient characteristics to fluctuations in physician behavior.

Sources of Physician Capacity As discussed in Section 1.3, physician capacity is computed based on the number of scheduled operations and operation days within seven days of the focal initial office encounter. Combining these measures into a single capacity indicator addresses the issue that operations and operation days are not very informative in isolation, because their impact on available physician capacity very much depends on operation days and operations, respectively. Nevertheless, I test the robustness of my findings by exploring whether scheduled operations and scheduled operation days individually and jointly predict operation rates.

Table 1.5 presents the corresponding estimates, reporting the impact of scheduled operations and scheduled operation days on operation rates within 90 days of an initial office appointment. Columns 1, 2, 4, and 5 show that scheduled operation days and operations have the expected impact on operation rates, but are not individually significant. Columns 3 and 6 document that both scheduled operations and operation days are highly significant predictors of operation rates when they are combined in the same regression. Not surprisingly, the impact of operation days and operations and is captured much more precisely once accounting for both of these measures.

In theory, an additional operation day should increase physician capacity by 3.52 — the average number of operations per operation day — as much as an additional operation reduces it. I test this hypothesis based on the estimates obtained in Column 6 with the following Wald test:

$$\text{Scheduled Operation Days} + 3.52 * \text{Scheduled Operations} = 0 \tag{1.4}$$

The resulting F-statistic is 0.94 with a corresponding probability of 0.33, implying that I cannot reject the hypothesis. Thus, using the average number of operations per operation day to construct physician capacity appears reasonable.

1.6 Conclusions

In this chapter, I study the impact of available physician capacity on operations within a month, a quarter, and a year of the initial encounter between a patient and an orthopedic surgeon. I find that operation rates across all time periods are significantly higher when the physician has ample spare capacity to schedule an operation within seven days of the focal encounter. As my analysis accounts for physician-month fixed effects, this finding is unlikely to be related to differences in physician characteristics. Meanwhile, the peculiarities of the setting of this chapter suggest that physician capacity is exogenous to unobservable patient characteristics.

The evidence presented in this chapter suggests that a fairly substantial number of operative decisions in orthopedic surgery is made under consideration of the physician capacity available within a week of the initial encounter. The impact of these operative decisions on health outcomes is unknown. Nevertheless, there is no scenario in which the operative decisions documented in this chapter are optimal. If operative decisions made under consideration of physician capacity are beneficial, orthopedic surgeons should perform surgeries on comparable patients arriving during periods when physician capacity is more limited. If, on the other hand, health outcomes are unaffected — or even deteriorated — by surgical decisions made under consideration of physician capacity, non-operative management would constitute a better treatment option for the corresponding patients.

Changes to physician and patient behavior to eliminate the impact of physician capacity on operative decisions would not only be advantageous in theory, but also feasible in practice. The patients considered in this chapter schedule an office appointment rather than reporting to the emergency room, suggesting that they do not require immediate surgery. If surgery was beneficial, orthopedic surgeons could schedule these patients for surgery at a later date or refer them to a colleague. I observe neither response in my data. Meanwhile, if non-operative management constitutes an adequate treatment option, it is almost always more time- and cost-efficient that surgical treatment.

It is important to note that the impact of physician capacity on operative decisions does not solely depend on orthopedic surgeons. Patients may also alter their operative decisions conditional on the waiting time for a surgery. Perhaps, physicians with limited short-term capacity suggest their patients to think about their preference between conservative and surgical treatment for a couple of weeks. During this period, patients may learn to cope with their pain, making non-operative treatment feasible in the long run. Anecdotal evidence suggests that at least some patients behave according to this hypothesis, suggesting that — contrary to intuition — constrained physician capacity can be beneficial in some situations.

This chapter supplements an enormous body of research reporting abundant variation of operative decisions across regions. To the extent that physician capacity fluctuates across regions, the variation in operative decisions documented in this chapter could help to explain regional variation in operation rates. Unfortunately, reliable data on short-term physician capacity across regions is not readily available. Besides painting another picture of variation of surgical decisions, this chapter adds to the literature by addressing two of its concerns. First, it exploits variation of the same orthopedic surgeon over time, convincingly controlling for differences in physician characteristics. Second, it considers every patients seen by a physician within the context of an initial office encounter. By contrast, previous research typically considers patients with a specific diagnosis, making sample selection susceptible to diagnosis behavior (Song et al. 2010, Welch et al. 2011).

Variation in elective operative decisions is frequently related to the "gray area" of medicine (Chandra et al. 2011), where physicians exert substantial discretion because the benefits of surgical treatment are unclear. Additional evidence on the effectiveness of surgical and non-operative treatments in this area of medicine would be enormously helpful. The relationship between physician capacity and operation rates documented in this chapter could support this effort, because it suggests that physician capacity introduces exogenous variation into operative decisions. What is currently missing, however, is data that is tracking health outcomes across patients undergoing surgery and patients who are treated non-operatively. Putting these pieces together is a promising avenue for future research.

Chapter 2

Operative Delays and Hip Fracture Outcomes: Operation Room Availability as Instrumental Variable[1]

2.1 Introduction

Annually, about 1.0 percent of women and 0.4 percent of men aged 65 and older suffer from a hip fracture in the United States (Brauer et al. 2009). Understanding the relationship between operative delays and outcomes experienced by these patients is crucial to determine the importance of treating these patients as soon as possible after their admission.

The evidence on the impact of operative delays on patient outcomes continues to be inconclusive. Holmberg et al. (1987) do not find adverse effects of operative delays up to one week after admission. By contrast, more recent evidence suggests that operative delays are correlated with worse patient outcomes.[2] In a systematic review, Khan et al. (2009) conclude that surgery within 48 hours of admission reduces length of stay and may reduce complications and mortality.

[1]Co-authored with David Ring and Mark S. Vrahas.

[2]See for instance Fox et al. (1994), Rogers et al. (1995), Weller et al. (2005), Moran et al. (2005), and Bottle and Aylin (2006).

Estimating the relationship between operative delays and the outcomes of patients after hip fractures is complicated by differences in patient characteristics. Typically, severe cases require more time to be prepared for surgery while also featuring worse health outcomes. Unless patient characteristics simultaneously influencing operative delay and patient outcomes are observed perfectly — which we believe to be impossible — estimates of the impact of operative delays on patient outcomes are biased.

We address this issue by exploiting quasi-random variation in the availability of operation rooms at the time of patient arrival. While the operation room capacity available to trauma surgeons — who treat the vast majority of hip fractures in our setting — has a small but statistically significant impact on operative delay, it presumably does not correlate with patient characteristics and outcomes. Thus, instrumenting operative delay with operation room availability allows for causal inference.

2.2 Methods

2.2.1 Data

We analyze data provided by the Research Patient Data Registry at Massachusetts General Hospital, an academic center located in Boston, Massachusetts. We consider patients seeing an orthopedic surgeon at the hospital between 2003–2012. Patients are included if they are older than 59 and have been diagnosed with a hip fracture based on the following ICD-9 codes: 820.00, 820.01, 820.02, 820.03, 820.09, 820.10, 820.11, 820.12, 820.13, 820.19, 820.20, 820.21, 820.22, 820.30, 820.31, 820.32, 820.8, 820.9. We consider only a patient's initial hip fracture diagnosis and focus on patients who first reported to the emergency room and were subsequently operated at the same hospital.[3]

Operative delay is defined as the time between the patient's arrival in the emergency room and the start of the operative procedure. We use operative delay as a continuous measure captured in hours. We also consider binary indicators of operative delay depending

[3]We exclude patients without an applicable emergency discharge note or operation record.

on whether a patient was operated with 24 or 48 hours of admission to the hospital. Patients waiting more than 240 hours for their operation — one percent of the sample — are excluded. Subsequently, our sample contains 1,977 patients.

We consider five outcome measures. Death is defined within 30 days of the operation date. Readmission is defined within 30 days of the discharge date if it resulted in another inpatient stay. Reoperation is defined within 90 days of the discharge date. This indicator variable is based on the following CPT codes: 10140, 10160, 10180, 11000, 11001, 11043, 11044, 11047, 26990, 26991, 26992, 76942, 77012, 77021, 97597. Infection accounts for both surgical site infections and more general infections such as pneumonia, sepsis, and urinary tract infections within 90 days of the discharge date. Finally, transfusion indicates whether the patient received a transfusion within 24 hours after the end of the surgical procedure.

We adjust for a broad range of patient characteristics. Besides gender and five-year age-groups[4] covariates include the type of hip fracture diagnosis specified as fixed effects.[5] We use two types of indicators to capture a patient's overall health condition. First, we use each of the categories of diagnoses used in the Charlson comorbidity index (Charlson et al. 1987, Quan et al. 2011) as fixed effects.[6] In addition, we create a variable indicating whether the patient reported to the hospital within 30 days before the admission date.

Finally, we account for the time of patient arrival at the emergency department. We control for fixed effects for the hour of the day, day of the week, and year. The first two covariates are meant to capture variation in resource availability during the day and the week. They also account for potential variations of patient characteristics across hours or weekdays. By contrast, year fixed effects adjust for long-term time trends in the treatment of hip fractures.

[4]Patients aged 95 and older are grouped in the final age group.

[5]ICD-9 codes 820.00, 820.09, 820.20, 820.21, 820.22, and 820.8 are grouped individually. The remaining ICD-9 codes account for less than three percent of the patients in our sample and are grouped together as residual category.

[6]The following categories are included in the Charlson comorbidity index: Chronic pulmonary disease, rheumatologic disease, chronic diabetes, renal disease, congestive heartfailure, dementia, mild liver disease, hemiplegia, paraplegia, malignancy, moderate and severe liver disease, AIDS, and tumor.

2.2.2 Operation Room Availability

Operation room availability is measured using records tracking when an operation is scheduled, performed, or cancelled by any of the attending trauma surgeons practicing at the hospital. In our setting, these surgeons conduct the majority of hip fracture operations.[7] We approximate the trauma surgeons' access to operation room time by considering the operation room time that has already been requested at the time of a patient's arrival and the number of operations previously scheduled during that time.

We do not observe the operation room time that has been requested by the trauma surgeons before a patient's arrival. To address this issue, we estimate the requested operation room time by counting the number of days during which a trauma surgeon had scheduled at least one operation prior to a patient's arrival in the emergency room. Subsequently, we compute the surgeon's long-term average capacity by dividing his number of operations by the number of operation days over the entire sample period.[8] We multiply the resulting average number of operations per operation day by the number of operation days scheduled by that surgeon, resulting in the total operation room capacity of that surgeon. We proceed analogously for each trauma surgeon and sum the operation room capacity across all of the attending trauma surgeons.

To obtain the available operation room capacity, we subtract the number of operations scheduled before a patient's arrival — reflecting the unavailable capacity — from the total operation room capacity computed above. Thus, our measure captures how easily surgeons can obtain additional operation room time at the time of patient arrival, although it is by no means perfect. We consider operations to be performed within four days after a patient's arrival. This time interval covers the operative delays of almost 95 percent of the selected patients.

---

[7] Besides trauma surgeons, patients with hip fractures are sometimes operated on by fellows and by attending orthopedic surgeons specializing in elective hip replacement rather than trauma surgery.

[8] For this computation we consider all of the operations performed by the corresponding trauma surgeon.

2.2.3 Estimation

We strive to estimate the impact of operative delays on patient outcomes based on the following regression model:

$$O_i = \alpha + \beta D_i + \gamma \mathbf{X}_i + \delta \mathbf{P}_{j(i)} + \theta \mathbf{T}_{t(i)} + \epsilon_i \tag{2.1}$$

Here, O denotes the outcome of patient i. D denotes the operative delay faced by patient i. $\mathbf{X}$ comprises a vector adjusting for patient characteristics described in more detail in Section 2.2.1. Meanwhile, $\mathbf{P}$ includes fixed effects for physician j operating on patient i.[9] $\mathbf{T}$ controls for the hour of the day, day of the week, and year, all specified as fixed effects. Finally, ϵ denotes the patient-specific error term. We estimate ordinary least square models despite the binary character of outcomes, because subsequent instrumental variable estimation is much more robust within this framework compared to a logistic or probit regression model.

The identifying assumption underlying (2.1) is that the covariates are uncorrelated with unobserved determinants of patient outcomes, which are captured by the error term ϵ. An enormous body of research — summarized by Shiga et al. (2008) — points to the possibility that unobserved patient characteristics may impact both operative delays and patient outcomes. If this was the case, coefficients obtained after estimating (2.1) would be biased.

Extensively controlling for patient characteristics clearly reduces the bias arising from differences in patient characteristics.[10] Nonetheless, it appears close to impossible to adjust for patient characteristics to an extent that unobservable patient characteristics can be reasonably expected to be uncorrelated with patient outcomes. Rather than seeking to increase the preciseness of patient covariates we suggest ensuring the validity of the identifying assumption underlying (2.1) by using an instrumental variable.

[9] Surgeons performing fewer than 50 surgeries in our sample are assigned to a residual category.

[10] For instance, Bottle and Aylin (2006) analyze the impact of at least one day of operative delay on death in hospital for patients suffering from a hip fracture. They find that controlling for patient characteristics reduces the odds ratio of operative delay 1.39 to 1.27.

2.2.4 Instrumental Variables

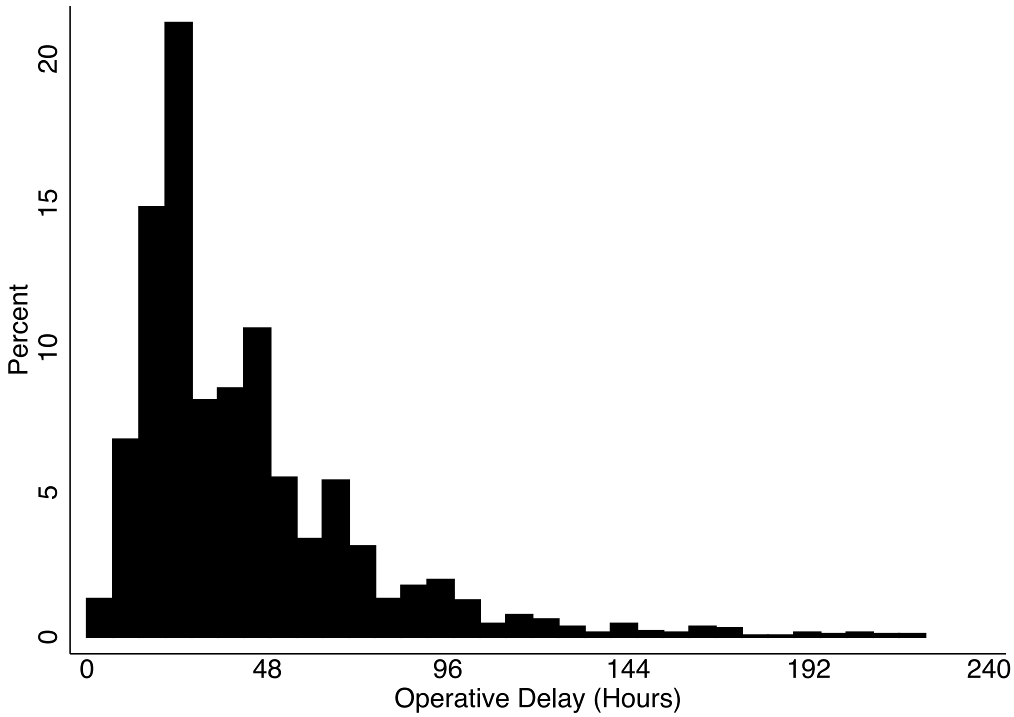
Instrumental variables were first introduced by Wright (1928) and have become widely used by economists and statisticians. Applications in the medical literature include McClellan et al. (1994). This concept addresses issues arising from the correlation between covariates and the error term in a regression model. To be valid, an instrumental variable needs to be both correlated with the covariate of interest and uncorrelated with the error term conditional on the other covariates. Consequently, an instrumental variable fixing (2.1) needs to predict operative delay without otherwise being related to the patient's outcome. We consider operation room availability as such an instrumental variable, hypothesizing that operation room availability has an influence on operative delay without being otherwise correlated with patient characteristics.

2.3 Results

2.3.1 Study Population

Figure 2.1 presents the distribution of operative delay across the study population. About a third of patients are treated within 24 hours after being admitted to the emergency room. 72 percent of patients are treated within 48 hours. Figure 2.2 plots average patient outcomes across categories of operative delay. A positive correlation between operative delay and adverse patient outcomes is evident. For instance, mortality amounts to 3.1 percent for patients being operated within 24 hours of admission. By contrast, patients who wait more than 96 hours for their operation die within 30 days in 10.8 percent of cases. We strive to explain to which extent this correlation is causal.

Table 2.1 summarizes the data used in this study by grouping patients according to the operation room availability available at the time of their arrival in the emergency room. We consider quintiles of operation room availability. Operation room availability ranges from -2.8 to 4.0, implying that the number of available surgery slots within four days



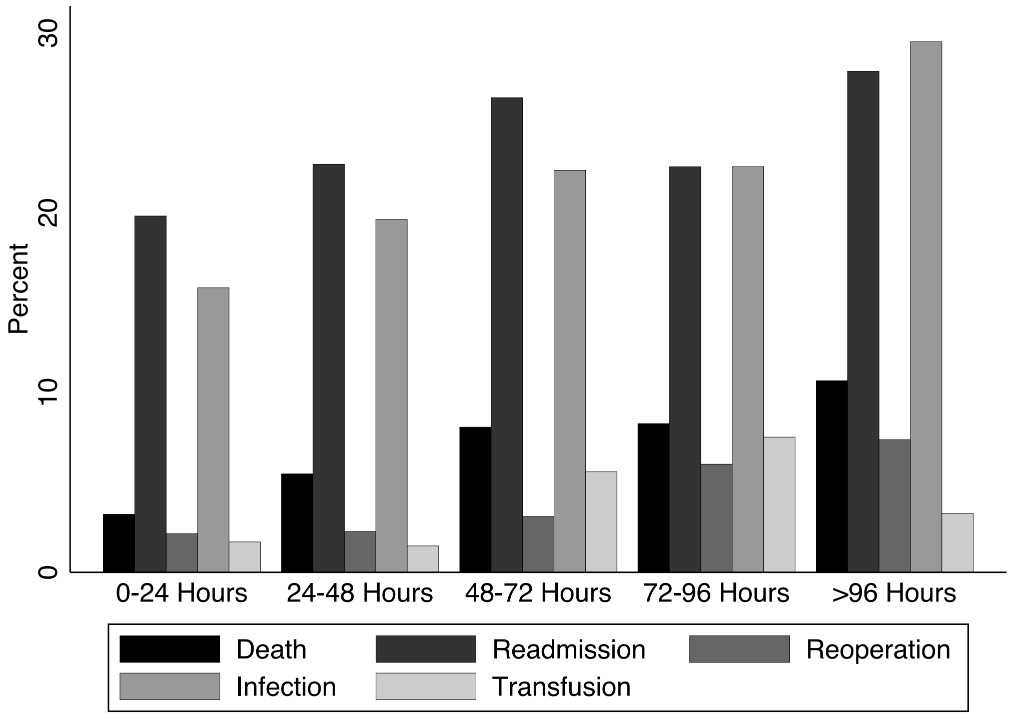
Notes: This figure shows the distribution of operative delays across the sample of patients analyzed in this study.

Figure 2.1: *Operative Delays*

of patient arrival varies substantially.[11] Operative delay steadily declines with operation room availability. Compared to the bottom quintile of operation room availability, the top quintile features operative delays that are almost five hours shorter on average. This effect is statistically significant at the five percent level.

Interestingly, the bottom quintiles of operation room availability features not only the longest operative delays, but also the worst average patient outcomes. The mortality rate in the bottom quintile is roughly two percentage points higher than in any other quintile. Similarly, the values for readmission, reoperation, and infection are highest in one of the bottom quintiles. These effects are not statistically significant, however.

[11]Operation room availability is negative if physicians have already scheduled more operations than their average capacity multiplied by the number of operation days.



Notes: This figure shows the means of patient outcomes for the given categories of operative delay.

Figure 2.2: *Operative Delays and Patient Outcomes*

Patients are indistinguishable across quintiles of operation room availability, confirming our identification strategy. By contrast, operation room availability varies significantly with the hour and day of patient arrival. Patients arriving later during the day are exposed to less operation room availability. Given our definition of operation room availability this makes perfect sense. As the day progresses, trauma surgeons schedule more operations. At first sight it may appear somewhat surprising that patients arriving earlier in the week are exposed to less operation room availability. It should be noted, however, that patients arriving during the weekend benefit from ample operation room availability during subsequent days than patients arriving before the weekend.[12]

[12]We define operation room availability within four days of a patient's arrival, measuring it during during four weekdays for patients arriving on Sundays or Mondays, during three weekdays for patients arriving on Tuesdays or Saturdays, and during two weekdays for patients arriving on Wednesdays, Thursdays, or Fridays.

Table 2.1: *Summary Statistics by Operation Room Availability Quintile*

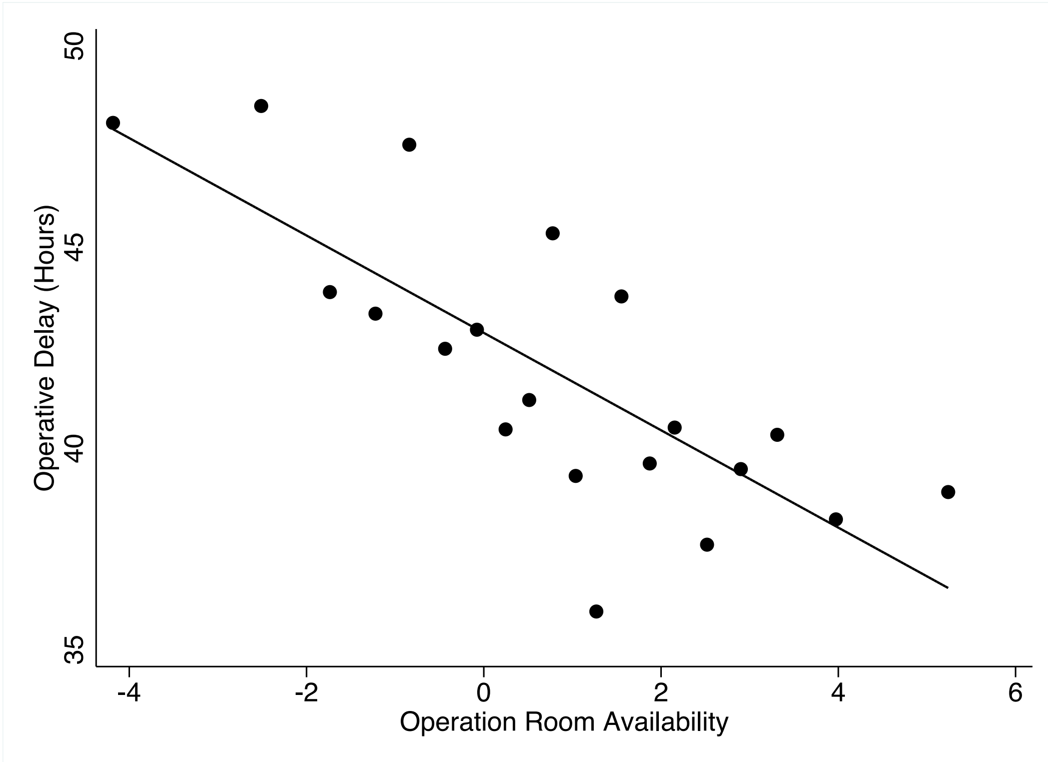
	Operation Room Availability Quintile				
	1	2	3	4	5
Operation Room Availability	-2.822	-0.2958*	0.973*	2.111*	3.957*
Operative Delay	45.11	41.82	41.42	40.83	40.32*
Outcomes					
Death (30 Days)	0.074	0.057	0.050	0.055	0.048
Readmission (30 Days)	0.239	0.250	0.196	0.229	0.220
Reoperation (90 Days)	0.031	0.036	0.030	0.023	0.027
Infection (90 Days)	0.191	0.224	0.201	0.182	0.184
Transfusion (24 Hours)	0.018	0.034	0.038	0.026	0.022
Patient Characteristics					
Age	79.60	80.26	79.72	80.36	79.33
Percent Male	0.308	0.320	0.304	0.281	0.298
Percent White	0.964	0.923	0.922	0.953	0.937
Charlson	3.112	2.835	2.972	2.662	3.080
Timing					
Hour Of Day	14.46	14.01	13.02*	12.20*	12.36*
Day Of Week	3.303	3.887*	4.339*	4.231*	4.690*
Observations	393	388	398	385	413

*Notes: This table shows summary statistics by quintile of operation room availability across all attending trauma surgeons within four days of patient arrival. * denotes significant difference from the bottom quintile at the 5 percent level, computed using standard errors clustered by admission date.*

2.3.2 Operation Room Availability and Operative Delays

Figure 2.3 visualizes the relationship between operation room availability and operative delay, adjusting for hour and weekday fixed effects. Evidently, operative delays decrease with operation room availability. Operative delays average between 45 and 50 hours when operation room availability is low, but decrease to below 40 hours when operation room availability is high.

Table 2.2 investigates the impact of operation room availability on operative delay in more detail. In Column 1, we estimate this relationship without any covariates. The resulting coefficient suggests that each unit of operation room availability — corresponding to an already available surgery slot by one of the attending trauma surgeons within the next



Notes: This figure shows the correlation between of operation room availability and operative delays adjusting for hour and weekday fixed effects.

Figure 2.3: *Operation Room Availability and Operative Delays*

four days — reduces operative delays by about 46 minutes. In Column 2, we add hour and weekday fixed effects, which have a substantial impact on the relationship between operation room availability and operative delay, resulting in an estimate of 62 minutes. We add patient covariates in Column 3 and physician fixed effects in Column 4, resulting in estimates of approximately 69 minutes.

To put these results into perspective, we multiply the coefficient estimates obtained in Table 2.2 by 2.44, a one standard deviation increase in operation room availability. Based on the coefficient estimate reported in Column 4, a one standard deviation increase in operation room availability decreases operative delay by 2.8 hours or 6.8 percent of the average operative delay. More intuitively, an attending trauma surgeon conducts about three operations per operation day. Consequently, if an additional day of operation room capacity

Table 2.2: *Operation Room Availability and Operative Delays*

	Operative Delay			
	(1)	(2)	(3)	(4)
Operation Room Availability	-0.767**	-1.042***	-1.153***	-1.156***
	(0.318)	(0.342)	(0.337)	(0.336)
Time Effects	No	Yes	Yes	Yes
Patient Effects	No	No	Yes	Yes
Physician Effects	No	No	No	Yes
Observations	1977	1977	1977	1977
Adjusted R2	0.003	0.024	0.047	0.047
Mean Operative Delay	41.886	41.886	41.886	41.886

*Notes: This table reports the effect of operation room availability on operative delays. Models are estimated by OLS and standard errors are clustered by admission date. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

was available at the time of patient arrival, the corresponding operative delay would be expected to be about 8 percent lower.

The most important finding documented in Table 2.2 is that the effect of operation room availability on operative delays is highly significant. We focus on the coefficient estimate reported in Column 4, because here we include all of the covariates used in our subsequent analysis. The T-statistic in Column 4 — obtained by dividing the coefficient estimate by its standard error — is 3.44. We compute the corresponding F-statistic by squaring the T-statistic, resulting in an estimate of 11.83. Importantly, this value exceeds 10, the value recommended as threshold for separating weak from strong instruments (Staiger and Stock 1997, Stock et al. 2002). We conclude that operation room availability is a strong enough instrumental variable.

Besides being correlated strongly with operative delays, operation room availability must not be otherwise correlated with patient outcomes — at least conditional on covariates — to be a valid instrumental variable. If patient characteristics or provided care varied system with operation room availability, it would not only impact operative delays, but also patient outcomes.

Patients arriving with a hip fracture in the emergency room require immediate attention and in all likelihood do not time their arrival. We note, however, that patient characteristics might differ across hours of the day and day of the week. For instance, we expect to see more sport-related hip fractures during the weekend. To the extent that operation room availability also varies across hours and weekdays, this phenomenon might induce a systematic relationship between operation room availability and patient characteristics, violating the assumption that our instrumental variable is uncorrelated with the error term. To address this issue, we control for fixed effects of the hour of the day and the day of the week in all of our analyses. Conditional on these covariates, it appears unlikely that patient characteristics vary systematically with operation room availability in a way that would bias our results.

Even if patient characteristics did not vary with operation room availability, operation room availability might be still be correlated with the quality of provided care. If the quality of provided care subsequently influenced patient outcomes, the independence of the instrumental variable and the error term would be violated, implying that operation room availability is not a valid instrument. The most immediate concern is that patients may be operated by a fellow instead of an attending trauma surgeon during periods of low operation room availability. Consequently, we include indicators for the surgeon performing the operation. We believe that conditional on these covariates, provided care does not vary with operation room availability. Nonetheless, we note that the validity of our findings crucially depends on this assumption. Our results are biased to the extent that unobserved covariates vary with both operation room availability and provided care.

2.3.3 Operative Delays and Patient Outcomes

Table 2.3 reports our main results. In Panel A, we present estimates of the effect of operative delays on patient outcomes based on regular ordinary least squares (OLS) models. In Panel B, we show estimates of the same effect, this time using instrumental variable (IV) models exploiting the relationship between operation room availability and operative delays. All

Table 2.3: *Operative Delays and Patient Outcomes*

	Death	Readmission	Reoperation	Infection	Transfusion
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Operative Delay	0.063***	0.068**	0.034**	0.087***	0.036**
	(0.021)	(0.030)	(0.014)	(0.032)	(0.014)
Adjusted R2	0.017	0.083	0.029	0.053	0.005
Panel B: IV					
Operative Delay	0.182	-0.100	0.161	0.294	0.030
	(0.198)	(0.360)	(0.149)	(0.352)	(0.119)
Adjusted R2	-0.008	0.067	-0.026	0.026	0.004
First Stage F-Statistic	11.834	11.834	11.834	11.834	11.834
Observations	1977	1977	1977	1977	1977
Mean Outcome	5.665	22.661	2.933	19.626	2.731

*Notes: This table reports estimates of the impact of operative delays on patient outcomes. Coefficients are multiplied by 100 and can be interpreted as percentages. Models control for time, patient, and physician effects as described in Section 2.2.1. Standard errors are clustered by admission date. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

models account for patient characteristics described in Section 2.2.1 as well as physician, hour, weekday, and year fixed effects. All of the coefficients are multiplied by 100 and can be interpreted as percentages.

Panel A shows that there is substantial and statistically significant correlation between operative delays and all patient outcomes considered in this study. According to Column 1, an additional hour of operative delay increases the mortality rate within 30 days of the operation by 0.06 percent. Consequently, delaying the operation by one more day is expected to increase death by 1.51 percent. Relative to the sample mortality rate of 5.67 percent, an additional day of operative delay increases the mortality odds by 26.63 percent. This estimate is broadly in line with estimates previously reported in the literature.

The effects of operative delay on reoperation and transfusion are similar to its effect on death. Meanwhile, the odds of readmission and infection are less — but still significantly —

affected by operative delay. Delaying the operation of a patient by an additional day increases the probability of readmission by 1.63 percent, reoperation by 0.82 percent, infection by 2.09 percent, and transfusion by 0.86 percent. The corresponding odds ratios relative to the sample mean are 7.20 percent for readmission, 27.82 percent for reoperation, 10.64 percent for infection, and 31.63 percent for transfusion.

Panel B presents the IV estimates. Estimates are generally insignificant. Taken at face value, the IV estimates — with the exception of readmission — tend be positive and at least as large as the OLS estimates. The impact of operative delay on death and infection is approximately three times larger using IV instead of OLS estimates. The IV estimates for reoperation are almost five times as large as the corresponding OLS estimate. Though imprecisely estimated, all of these estimates suggest that correlation between operative delay and patient outcome is not solely related to differences in patient characteristics. By contrast, readmission rates appear to *decrease* with operative delay according to Column 2 of Panel B. Besides statistical noise, one potential explanation for this finding could be that patients whose operation is delayed stay longer in the hospital and experience complications during this time, while patients who are treated earlier are also discharged earlier and need to be readmitted for early complications.

2.3.4 Robustness

The IV estimates presented in Table 2.3 are statistically insignificant, providing very limited information on the causal impact of operative delays on patient outcomes. One potential explanation for this phenomenon is that the relationship between operation room availability and operative delays — while statistically significant — is fairly small. As documented in Table 2.1, moving from the bottom to the top quintile of operation room availability observed in our data reduces the average operative delay by less than five hours. At the same time, outcomes are quite noisy as is evident both in Figure 2.2 and Table 2.1. Consequently, the IV estimates are based on small variation in operative delays and predict noisy variation of outcomes, leading to large standard errors.

In Table 2.4, we consider different specifications of operative delay, hoping that operation room availability predicting these measures of treatment delay more accurately. If this was the case, the resulting IV estimates should be more precise. We specify two variables indicating whether operative delay exceeds 24 or 48 hours, respectively.[13] Corresponding estimates are presented in Columns 2 and 3, while Column 1 repeats the baseline estimate previously shown in Column 1 of Table 2.3. In Table 2.4, we report estimates using death as outcome variable. We note that we obtain substantially similar results for the other outcomes considered in this study.

The first-stage *F*-statistic reported in Column 2 is substantially larger than the corresponding figure of the baseline specification shown in Column 1, suggesting that operation room availability predicts operative delay of more than 24 hours more precisely than overall operative delay. Unfortunately, this improved first-stage regression does not translate to more precise estimates. In both Columns 1 and 2, the IV estimates of operative delay on death roughly triple relative to the corresponding OLS estimates but at the same time become statistically insignificant. Meanwhile, specifying operative delay exceeding 48 hours as explanatory variable — reported in Column 3 — results in a first-stage *F*-statistic that is just half as big as the *F*-statistic corresponding to the OLS specification.

Finally, we consider whether outliers have an impact on our results in Column 4. In this specification, we only consider patients operated within 96 hours of admission, excluding about 5 percent of patients that had to wait longest for their operation. Relative to the baseline estimate in Column 1, the OLS estimate is slightly larger, while the IV estimate is somewhat smaller. The ratio of IV to the OLS estimate drops from almost three to slightly more than two, unsurprisingly suggesting that our estimates are quite sensitive to sample specification.

[13] 64 and 28 percent of patients are not treated within 24 and 48 hours, respectively.

Table 2.4: *Operative Delays and Death*

	Death			
	(1)	(2)	(3)	(4)
Panel A: OLS				
Operative Delay	0.063***			0.073***
	(0.021)			(0.028)
Operative Delay $>$ 24 Hours		3.727***		
		(1.025)		
Operative Delay $>$ 48 Hours			3.948***	
			(1.346)	
Adjusted R2	0.017	0.016	0.016	0.016
Panel B: IV				
Operative Delay	0.182			0.141
	(0.198)			(0.242)
Operative Delay $>$ 24 Hours		10.324		
		(11.243)		
Operative Delay $>$ 48 Hours			17.167	
			(19.278)	
Adjusted R2	-0.008	-0.002	N/A	0.013
First Stage F-Statistic	11.834	17.389	6.101	15.366
Observations	1977	1977	1977	1857
Mean Outcome	5.665	5.665	5.665	5.331

*Notes: This table reports estimates of the impact of operative delays on death within 30 days of the operation. Coefficients are multiplied by 100 and can be interpreted as percentages. Models control for time, patient, and physician effects as described in Section 2.2.1. Standard errors are clustered by admission date. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

2.4 Discussion

We find that operative delays are substantially and significantly correlated with higher mortality, readmission, reoperation, infection, and transfusion rates after hip fractures. The causal effects underlying this relationship remain unclear. Using operation room availability as instrumental variable allows for causal interpretation, but results in noisy estimates.

2.4.1 Implications

Understanding the relationship between operative delay and patient outcomes is crucial to assess to which extent operations may be delayed. While we are unable to determine the extent to which delaying treatment causally deteriorates patient outcomes, our findings generally do not suggest that operative delays matter less than suggested by their correlation with patient outcomes. Contrary to our expectation, unobservable differences in patient characteristics — at least according to our very imprecise findings — do not seem to account for the bulk of the relationship between operative delays and patient outcomes. If these findings were to be confirmed in additional studies, the implications are clear: Operative delays should be avoided with urgency and their impact on patient outcomes cannot simply be disregarded as related to case severity.

2.4.2 Limitations

Several limitations of this study deserve further discussion. First and foremost, operation room availability constitutes an instrumental variable that may have both theoretical and practical issues. From a theoretical perspective, it remains debatable whether the availability of operation rooms is truly independent of patient outcomes besides its impact on operative delays. Another concern is whether we capture operation room availability precisely enough. In practice, the first-stage relationship between operation room availability and operative delays is significant but fairly small, presumably leading to the enormous standard errors we observe when estimating the effect of operative delays on outcomes.

A second important limitation concerns the specific circumstances of our setting. As a major teaching hospital, Massachusetts General Hospital disposes over a large number of orthopedic surgeons. Consequently, major delays are very rare and almost always related to medical reasons rather than operation room and surgeon availability. Consequently, about 80 percent of patients are treated within the two days following their admission. The resulting variation in operative delays that is not related to patient characteristics is likely smaller than in a hospital with more constrained resources, complicating identification.

Finally, some data issues deserve being mentioned. We observe the time of a patient's arrival in the emergency room, but do not know the initial fracture time. This unobservable variable may affect patient outcomes. However, as long as initial fracture time is unrelated to operation room availability — which we believe to be the case — it would not bias instrumental variable estimates. Meanwhile, some of the patient outcomes — readmission, reoperation, and infection — are only tracked within the hospital and are only measurable if the patient is still alive, potentially biasing our estimates.

Chapter 3

Concentrated and Broad Team Familiarity: Evidence from Orthopedic Surgery[1]

3.1 Introduction

Numerous organizations assemble professionals in teams to complete a single project (Edmondson and Nembhard 2009, Huckman and Staats 2011). Subsequently, these teams are disassembled and their members are assigned to new projects. Arrow and McGrath (1995) have characterized these teams as *fluid*. The benefits of fluid teams are typically understood in terms of increased flexibility, enabling organizations to simplify the assignment of projects to individuals. Relying on fluid teams, however, implies that team members are less familiar with each other than in traditional settings where teams remain intact across projects.

Studying the familiarity of fluid teams is complicated by the substantial variation of their familiarity across team members and over time. The usual measure of team familiarity — the number of past collaborations — frequently fluctuates within fluid teams. Some members of fluid teams may have frequently collaborated in the past, while other members may have never worked together before. In addition, fluid teams collaborate in irregular intervals. Consequently, their familiarity varies with the number of recent collaborations to the extent that team familiarity decays. By contrast, traditional teams typically keep

[1] Co-authored with Robert S. Huckman and David Ring.

working together across projects, suggesting that team familiarity neither varies across team members nor decays.

We address the varying degrees of familiarity within a fluid team by distinguishing between *concentrated* and *broad* team familiarity. In its extreme form, concentrated team familiarity stems from past collaboration between two team members involved in the focal project. In contrast, the broadest measure of team familiarity captures prior collaborations involving *all* members of the focal team. Developing broad familiarity is often cited as a benefit of creating dedicated teams. As we see in our context, the breadth of familiarity can be theoretically viewed as a continuum running between these extremes. Separating a team's past projects based on the number of members involved in both the past and the focal project enables us to account for variations in familiarity within fluid teams. Subsequently, we analyze how concentrated and broad team familiarity affect team performance over different time periods, testing whether team familiarity decays.

Understanding how fluid teams develop familiarity is crucial for any organization weighing the costs and benefits of relying on such teams. The impact of familiarity on the performance of fluid teams has been acknowledged by theoretical contributions (Edmondson and Nembhard 2009) and studied in a number of empirical settings. Katz (1982) examines the impact on familiarity on communication patterns based on cross-sectional data. Edmondson et al. (2001) provide qualitative evidence that team familiarity and communication affect learning. This notion is confirmed by Reagans et al. (2005) and Huckman et al. (2009), who demonstrate that familiarity improves the performance of fluids teams in surgery and software services. Similarly, Huckman and Pisano (2006) find that individual performance is firm-specific and suggest that team familiarity might be an important determinant of these variations in cardiac surgery.

We extend this literature in two ways. First, distinguishing between concentrated and broad team familiarity adds an important dimension to the decision organizations face when relying on fluid teams. Presumably, less coordination — such as customized scheduling of staff — is required to increase concentrated team familiarity. That said, the benefits of a given amount of broad team familiarity might be greater than that created by the same amount of concentrated team familiarity. Similarly, organizations are confronted with a trade-off regarding the stability of fluid teams. Depending on the relative accumulation and decay of benefits from concentrated versus broad familiarity, it might be optimal to keep fluid teams together for extended periods of time.

Our empirical analyses focus on orthopedic surgeries conducted at a major academic medical center in the United States. In this setting, dozens of surgeons and hundreds of assisting surgeons, anesthesiologists, and nurses are assigned to fluid teams to perform surgeries. This process leads to substantial variation in team familiarity across the procedures in our sample. Mandatory operation reports enable us to track team assignments over several years and link them to procedure duration, a well-established performance measure of surgical procedures (Pisano et al. 2001, Edmondson et al. 2003, Reagans et al. 2005). Though this measure is primarily concerned with efficiency, it is also correlated with surgical site infections (Leong et al. 2006), perioperative mobility (Kessler et al. 2003), and survival (Ong et al. 2008) for certain types of procedures.

In this setting, team experience can be measured by the number of past operations performed jointly by team members collaborating on the focal operation. Subsequently, we distinguish past operations based on the number of team members associated with both the past and the focal procedure. We find that past operations reduce procedure duration more if they involve more of the focal team's members. Broad familiarity — developed in operations involving the surgeon and two or more other team members — reduces procedure duration by more than twice as much as concentrated familiarity developed in collaborations between the surgeon and one other team member. These effects are most pronounced for operations conducted on the same day before the focal operation. Operations conducted 1–10 days before the focal surgery have a way smaller, but otherwise similar and statistically significant effect on procedure duration. By contrast, procedure duration is not affected by any type of operations conducted more than 10 days before the focal operation. We conclude that — at least in our setting — the familiarity of a team fades quickly.

3.2 Familiarity of Fluid Teams

We study operations performed by orthopedic surgeons jointly with assisting surgeons, anesthesiologists, and nurses. These teams are assigned based on the team members' schedules. Within this setting, teams are fluid and exhibit varying degrees of familiarity within the same team as well as enormous fluctuation in the number of recent collaborations. That is, a surgeon may have collaborated frequently with the assisting surgeon but not with the anesthesiologist. Meanwhile, a surgeon may have collaborated with the assisting surgeon frequently in the past week in some instances and never in the past week in other circumstances. Accurately measuring a team's familiarity in this context is challenging.

Numerous studies measure the familiarity of a team by counting the number of projects its members conducted jointly during a given time period. In traditional settings, this measure captures team experience reasonably well. In these settings, teams usually involve the same members over time, implying a standard learning curve accurately reflected by the number of past projects or units completed. By contrast, measuring the familiarity of fluid teams is complicated by the two issues we anecdotally described in the previous paragraph. First, the number of past cooperations can vary substantially across pairwise combinations of individuals within a team. Second, to the extent team familiarity decays, it fluctuates constantly with the number of recent collaborations. In this paper, we investigate the empirical importance of both of these factors.

3.2.1 Measuring Team Familiarity

Fluid teams frequently feature varying degrees of pairwise familiarity between various combinations of two members of the same team. Some pairs of team members may have collaborated very often in the past, while other members may barely know each other. Across numerous settings, shared experience developed among some, but not all, team members constitutes a major determinant of overall team familiarity. Consequently, the familiarity of fluid teams is not accurately captured by the number of previous cooperations involving *all* team members.

Previous research defines team familiarity based on the number of collaborations among pairs of team members. Capturing team familiarity at the level of a pair — the smallest unit of a team — is intuitively appealing, because it accounts for any potential source of team familiarity. Pairwise collaborations are aggregated across the whole team to compute its familiarity. The aggregation of pairwise collaborations implicitly defines the importance of each pairwise collaboration for the overall familiarity of the team. Typically, each pairwise collaboration is weighted equally. We analyze the implications of this assumption.

Reagans et al. (2005) and Huckman et al. (2009) define team familiarity as follows:

$$\text{Team Familiarity} = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{P_{ij}}{\frac{N(N-1)}{2}} \tag{3.1}$$

P denotes the number of past projects involving both team member i and team member j belonging to the N team members performing the focal project. Team familiarity is computed by summing P across all N team members, implying that all pairwise collaborations are weighted equally. Subsequently, the sum of pairwise collaborations is divided by $\frac{N(N-1)}{2}$ to compare the familiarity of teams including a varying number of team members. We ignore this adjustment because we focus on measuring the familiarity of teams that are composed of the same number of members.

The plausibility of (3.1) as a measure of team familiarity relies on the assumption that each of the pairwise collaborations contributes to shared experience in a similar way. That is, a three-member team in which all three members have worked together exactly once before would be assumed to have the same level of team familiarity as a three-member team where one pair of members has worked together twice and another pair has worked together once and the third pair has not worked together at all. This assumption yields at least three testable implications. First, it implies that a pairwise collaboration of any two team members affects team familiarity in the same way regardless of their importance to the team. Second, team familiarity increases linearly with the number of pairwise collaborations. Finally, team familiarity does not depend on whether pairwise collaborations were accumulated across different projects or within a single project. We test the plausibility of the final implication.

In many teams — including the surgical teams we study — certain members may play leadership roles that make their familiarity with other team members particularly important in determining performance. In the case of orthopedic surgery, one can measurably claim that the surgeon exerts a greater influence on patient outcomes than the other members of a surgical team. That is, assuming that pairwise collaborations of *any* two team members impact team familiarity equally appears implausible in this setting. We address this issue by measuring familiarity only using past operations that included the surgeon. Subsequently, we test whether pairwise collaborations of the surgeon and any other team member have a comparable influence on team performance.

3.2.2 Concentrated and Broad Team Familiarity

According to the traditional measure defined in (3.1), team familiarity does not account for disproportional benefits arising from the collaboration of more than two of the focal team members. Consider a surgical team consisting of a surgeon, an assisting surgeon, and an anesthesiologist. An operation involving all three team members would increase team familiarity by two units.[2] The same increase in team familiarity is achieved by two operations involving the surgeon and one of the other team members.

We investigate this question by distinguishing previous cases based on the number of involved team members, separating concentrated and broad team familiarity. We define concentrated team familiarity as the number of past collaborations conducted jointly by two — but no more — of the focal team members. Broad team familiarity is defined as the number of past collaborations involving three or more of the focal team members. Based on prior work regarding team familiarity, we would expect both concentrated and broad team familiarity to improve performance. As broad familiarity is built on more pairwise collaborations than concentrated familiarity, we expect that:

Hypothesis 1. *Broad team familiarity improves team performance more than concentrated team familiarity.*

[2] Pairwise collaborations not involving the surgeon are ignored as the surgeon is the central team member.

Put differently, we hypothesize that the benefits of concentrated team familiarity — stemming from a single pairwise collaboration — are exceeded by the advantages of broad team familiarity stemming from multiple pairwise interactions. This hypothesis raises an obvious question: How much more does broad team familiarity improve team performance relative to concentrated team familiarity?

We compare concentrated and broad team familiarity based on the corresponding number of pairwise collaborations. For instance, a three-member team in which all members collaborated once would have accumulated one unit of broad familiarity and three units of pairwise collaboration. Meanwhile, a three-member team in which two members collaborated once would have developed one unit of concentrated familiarity and one unit of pairwise collaboration. Here, we wonder whether broad familiarity improves team performance by three times as much as concentrated familiarity.

Theoretically, the relative impact of concentrated and broad familiarity on team performance is ambiguous. Consider a three-member team in which two members collaborated in the past with another individual not involved in the focal project. Replacing this individual with the third focal team member may improve performance by less than the the three-fold increase in pairwise collaborations. On the other hand, familiarity developed jointly among most or all team members during the same project may have benefits exceeding the increase in pairwise collaborations.

Besides capturing a different type of pairwise collaboration — performed in bulk during the same project rather than in isolation across distinct projects — broad team familiarity differs from concentrated team familiarity because it ensures that familiarity across a team is balanced. By contrast, concentrated familiarity may be unbalanced if it is concentrated among some team members. As the traditional measure of team familiarity increases linearly with the number of pairwise collaborations, the balance of these collaborations across team members does not affect team familiarity. That is, a three-member team in which two team members collaborated twice in the past has the same concentrated familiarity as a three member-team in which one member collaborated with each of the other members once.

In our empirical investigation, we explore how team performance varies with its familiarity captured by (3.1), concentrated familiarity, and broad team familiarity. We distinguish between *pairwise* estimates according to (3.1) and *teamwise* estimates based on our concept of concentrated and broad team familiarity. Considering pairwise and teamwise estimates gives us a reference point for the comparison of concentrated and broad team familiarity. In addition, it enables to assess the plausibility of (3.1), testing whether the relationship between team familiarity implied by (3.1) and the number of team members involved in both the past and the focal project is supported empirically.[3]

We focus on the overall impact on concentrated and broad team familiarity and are unable to disentangle the various sources potentially accounting for differences between concentrated and broad team familiarity. It thus remains unclear whether our findings stem from decreasing or increasing returns to pairwise collaborations or from improved balance of familiarity implied by broad team familiarity.

3.2.3 Decay of Team Familiarity

Due to frequent changes in team composition, the familiarity of fluid teams constantly shifts. In some situations, a surgeon and an assisting surgeon may have collaborated 10 times in the past 10 days. In contrast, at a different point in time, the same surgeon and assisting surgeon may not have collaborated at all during the prior 10 days. We consider whether team performance fluctuates across these situations. Put differently, we ask whether team familiarity decays. This question becomes relevant as one moves away from traditional settings in which the core of a given team may stay intact across several projects. Prior work on learning has found that the benefits of individual or firm experience decay over time. Like experience, we would expect that the benefits of team familiarity may decay as well, leading us to hypothesize:

Hypothesis 2. *The benefits of both concentrated and broad team familiarity decay.*

[3] According to (3.1), a past project involving two, three, four, and five team members yield one, three, six, and 10 units of pairwise collaboration, respectively. Thus, the relative importance of concentrated and broad team familiarity is determined a priori.

We test this notion by comparing how team performance is affected by team familiarity stemming from the most recent past and team familiarity accrued in the more distant past. These time periods are mutually exclusive. Based on these measures, we estimate whether team familiarity developed some time ago impacts team performance even when we control for team familiarity developed in the recent past.

3.3 Setting and Data

3.3.1 Setting

Massachusetts General Hospital (MGH) is major academic medical center located in Boston, Massachusetts. The hospital provides care across all specialties of medicine and is designated as a Level I Trauma Center. As of 2013, physicians at the hospital admit approximately 48,000 inpatients, manage nearly 1,500,000 outpatient visits, and perform more than 38,000 operations on an annual basis.

Our analysis focuses on the Department of Orthopedic Surgery at MGH. This department accounts for about 8 percent of all patient discharges and typically performs more than 30 percent of the total number of operations at MGH. We observe more than 50 orthopedic surgeons performing a meaningful number of operations at MGH. They perform surgeries in collaboration with hundreds of assisting surgeons, anesthesiologist, and nurses, resulting in an enormous number of potential team assignments. Teams are very fluid in this setting, frequently assembled for just a single operation. Importantly, the assignment of teams depend mostly on the schedules and availability of personnel which are determined way in advance. Individual surgeons have very limited impact on selecting team members. Thus, it appears unlikely that team performance affects team assignment, a feature that is crucial for our identification strategy.

Besides constant and substantial variation in team composition, orthopedic operations lend themselves to the analysis of team performance for several other reasons. Team performance can be accurately tracked by procedure duration, a well-established performance measure (Pisano et al. 2001, Edmondson et al. 2003, Reagans et al. 2005). This measure provides substantial variation both within and across surgeons. Furthermore, procedure duration clearly does not only depend on the surgeon's performance, but also on the efficiency of collaboration across the whole team. We capture both team composition and procedure duration from mandatory operation reports.[4]

3.3.2 Data

Our data includes all operations performed at the Department of Orthopedics at MGH from 2008–2011.[5] We observe 45,662 operations performed by 202 surgeons. 52 surgeons performed more than 100 operations during the observation period. Besides the surgeon, these data report the assisting surgeons, anesthesiologists, and nurses associated with the operation. This feature enables us to track how often an individual worked together with another professional or group of professionals over various time periods, constituting our measure of team familiarity.

In addition, our data identifies the type of procedure based on ICD-9 codes, the time and duration of the procedure, and the operation room where the operation took place. We benefit from a large number of patient characteristics, comprising the patient's Charlson index — a compound measure of comorbidity severity — in addition to age, gender, race and marital status (Charlson et al. 1987, Quan et al. 2011). The Charlson index is computed based on all diagnoses ever given to a patient at MGH before the operation date.[6]

Orthopedic surgeries are conducted by a team of surgeons, anesthesiologists, and nurses. These teams vary in size and composition depending on the requirements of a procedure and the availability of personnel. We focus on operations involving a surgeon, an assisting

[4]Relying on procedure duration as outcome enables us to pool different types of orthopedic surgeries for empirical analysis. Other quality measures are not easily comparable across different procedure types.

[5]This data excludes operations on MGH employees.

[6]Primary ICD-9 codes are missing for 1,452 patients. Procedure duration is missing for 233 patients. The operation room is missing for 173 patients. The Charlson index is missing for 214 patients. Race is missing for 1,455 patients. Marital status is missing for 1,374 patients.

surgeon — who could be a resident, fellow, or board-certified surgeon — an anesthesiologist, a circulating nurse, and a scrub nurse. These five professionals constitute the core surgical team.[7] We focus on the first operation of each patient during the observation period.[8] Our final analysis sample consists of 25,180 operations.[9]

3.3.3 Fluidity of Surgical Teams

As we observe operations involving 202 surgeons, 476 assisting surgeons, 376 anesthesiologists, 346 circulating nurses, and 452 scrub nurses, we expect substantial variation in the composition of the core surgical teams. Considering the surgeon as the focal team member, we calculate how often the surgeon collaborated with a given assisting surgeon, anesthesiologist, or nurse based on all available data. These calculations define the share of operations a surgeon conducts with a given other team member.

Based on a surgeon's collaboration shares, we calculate the Herfindahl index — a commonly used measure for market concentration — of the respective surgeon:

$$H_s^{Anesthesiologist} = \sum_a \left(\frac{Operations_{s,a}}{Operations_s}\right)^2 \tag{3.2}$$

$H_s^{Anesthesiologist}$ denotes the Herfindahl assessing the collaboration intensity between anesthesiologists indexed by a and surgeon s. *Operations* denotes the number of operations by the subscripted team members. We compute a Herfindahl index of collaboration intensity across all surgeons by weighting individual surgeon indices by their number of operations:

$$H^{Anesthesiologist} = \sum_s H_s^{Anesthesiologist} * \frac{Operations_s}{Operations} \tag{3.3}$$

Here, *Operations* denotes the number of operations conducted by all surgeons during the observation period. We proceed analogously with assisting surgeons and nurses.

---

[7] 11,484 operations were not performed by a core surgical team. In more than 10,000 of these operations, no assisting surgeon was participating in the operation.

[8] 9,248 operations were not primary operations.

[9] We use all 45,662 operations to compute individual and team experience.

Table 3.1: *Surgeons' Collaboration Concentration*

	Assistant	Anesthesiologist	Scrub	Circulator
Herfindahl Index (Percent)	4.37	11.23	6.26	8.91

Notes: This table shows the Herfindahl index representing the concentration of the surgeons' collaboration with each type of team member represented in a surgical team. The indices are computed individually for each surgeon and subsequently averaged across surgeons. Finally, surgeons are weighted by operation frequency.

Table 3.1 present the various Herfindahl indices translated to percentages. None of the indices exceeds 15 percent, indicating that collaboration is fairly unconcentrated. The importance of this lack of concentration is addressed below in our discussion of team assignment. It should be noted that the collaboration intensity between surgeons and assisting surgeons is biased downwards by the fact that some of the residents and fellows were working at MGH only for a portion of the observation period. Residents work at MGH for a five-year period straight out of medical school, where the first year is the internship. After completing residency at MGH or another hospitals, surgeons may complete a one-year fellowship. Most surgeons leave MGH after completing their education, working at different hospitals or in private practice.

3.3.4 Assignment of Surgical Teams

Analyzing the relationship between team familiarity and team performance is typically complicated by the fact that teams do not form randomly. Rather, teams performing well might choose to work together more frequently, resulting in a spurious correlation between familiarity and performance. Addressing this issue requires variation in the assignment of teams that is both rich and exogenous.

In our empirical setting, teams are composed primarily based on the type of operation and on the availability of employees and facilities. We adjust for the first factor by controlling for the primary ICD-9 code of each operation. Meanwhile, the availability of employees and facilities is largely random. Employee schedules are typically fixed before operations are scheduled. A substantial number of operations constitute emergencies and cannot be scheduled at all. Surgeons have very limited impact on team assignment, reflected by the low collaboration intensity with other members of the surgical team reported in Table 3.1.

The assignment of surgical teams is fairly random across days, but tends to exhibit substantial autocorrelation within days. Once a team has been composed to perform an operation at the start of the day, it is much more likely to stay together for the rest of the day than pure randomness would suggest. We note that this decision is not driven by the team's performance, but rather by scheduling considerations. The exact duration of operations is hard to predict, implying that keeping a team together avoids idle time for all team members. Consequently, surgical teams are more familiar with each other during operations that are scheduled later in the day. We address this issue below.

3.4 Empirical Strategy

3.4.1 Measuring Team Familiarity

As discussed in Section 3.2, we follow a large literature measuring the familiarity of a team using the volume of previous collaborations. However, we distinguish past operations based on the number of team members participating in both the past and the focal operation. Based on this distinction, we are able to separately consider the impact of concentrated and broad familiarity on team performance.

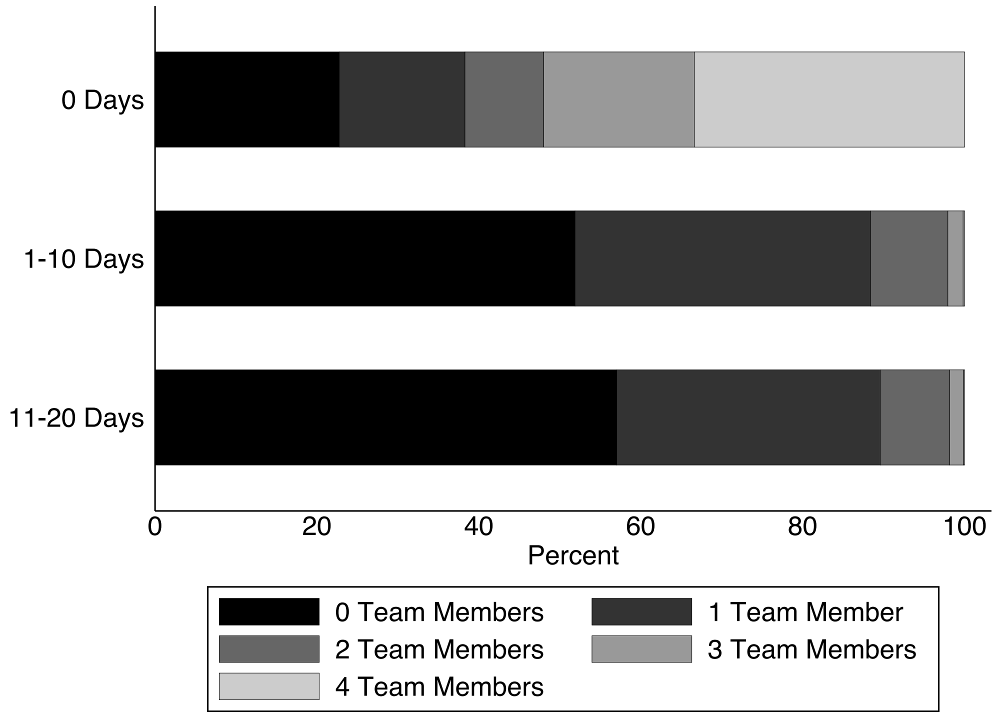
As noted previously, we consider the surgeon as the central team member. Thus, we ignore past operations not involving the surgeon when computing team familiarity. Though there is reason to expect the surgeon to be the central team member, we could alternatively specify concentrated and broad team familiarity by considering all team members as equally relevant. We report estimates based on both specifications and note that our results hold in either scenario. Besides the surgeon, a surgical team consists of four other team members discussed in Section 3.3.2. Consequently, we distinguish past operations performed by the surgeon based on whether one, two, three or four other team members collaborated with the surgeon in both the past and the focal operation.

Table 3.2: *Descriptive Statistics*

	Observations	Mean	Standard Deviation
Outcome			
Duration (Minutes)	25180	138.36	85.74
Surgeon			
Operations 0 Days	25180	1.56	1.80
Operations 1–10 Days	24932	9.47	7.08
Operations 11–20 Days	24748	9.79	7.33
Surgeon & 1 Team Member			
Operations 0 Days	25180	0.24	0.72
Operations 1–10 Days	24932	3.45	3.69
Operations 11–20 Days	24748	3.19	3.71
Surgeon & 2 Team Members			
Operations 0 Days	25180	0.15	0.55
Operations 1–10 Days	24932	0.90	1.84
Operations 11–20 Days	24748	0.84	1.86
Surgeon & 3 Team Members			
Operations 0 Days	25180	0.29	0.67
Operations 1–10 Days	24932	0.18	0.76
Operations 11–20 Days	24748	0.17	0.78
Surgeon & 4 Team Members			
Operations 0 Days	25180	0.52	0.86
Operations 1–10 Days	24932	0.02	0.23
Operations 11–20 Days	24748	0.01	0.20
Pair of Surgeon & Team Member			
Operations 0 Days	25180	3.49	4.17
Operations 1–10 Days	24932	5.87	6.74
Operations 11–20 Days	24748	5.42	6.84
Other Variables			
Experience (Years)	21350	14.97	9.01
Complexity (Work RVU)	20532	12.62	7.36

Notes: This table shows summary statistics for the key variables of our analyses.

Table 3.2 provides descriptive statistics for the most important variables analyzed in this paper. On average, an operation considered in this study takes about 138 minutes. This duration is defined by the time the patient enters and leaves the operation room. We report a surgeon's number of operations performed during the same day, within 1–10 days, and within 11–20 days before the focal operation. On average, a surgeon performs 1.56



Notes: This figure distinguishes a surgeon's prior operations on the same day, within 1–10 days, and within 11–20 days before the focal operation based on the number of team members that collaborated with the surgeon in both the past and the focal procedure.

Figure 3.1: *Prior Surgeon Operations and Focal Team Members*

operations on the same day before starting the focal operation. The corresponding averages within 1–10 days and 11–20 days of the focal operation are slightly less than 10 operations.

Figure 3.1 visualizes how a surgeon's operations are distributed across the number of team members participating both the past and the focal operation. On the day of the focal operation — denoted as 0 days — each operation type appears frequently. About 75 percent of operations involve at least one team member besides the surgeon. Every third operation is performed by all team members. By contrast, the percentage of operations involving three or four team members besides the surgeon is very small within 1–10 days and 11–20 days before the focal procedure. Thus, we focus on the day of the focal operation to analyze the effects of concentrated and broad team familiarity in detail. Subsequently, we group operations involving two or more additional team members together to examine the decay of team familiarity.

As discussed in Section 3.3.4, surgical teams are likely to collaborate in subsequent operations during the same day due to scheduling considerations. Consequently, teams performing operations later in the day are typically more familiar with each other. To the extent that surgeries scheduled later in the day are shorter — even when holding team familiarity constant — the observed relationship between team familiarity and procedure duration is biased. Surgeons tend to schedule easier procedures later in the day, making this a relevant concern. We address this issue by accounting for the number of operations that a surgeon performed previously during the same day using fixed effects.

3.4.2 Estimation

We use an OLS regression model to assess the relationship between team performance and different degrees of team familiarity:

$$\ln Duration_{ij} = \alpha + \beta \mathbf{OP}_j + \delta \mathbf{X}_i + \eta \mathbf{Y}_j + \theta \mathbf{T}_{ij} + \epsilon_{ij} \tag{3.4}$$

Here, observations are at the level of patient i undergoing an operation conducted by surgical team j including a surgeon, an assisting surgeon, an anesthesiologist, a circulating nurse, and a scrub nurse. *Duration* denotes the procedure duration in minutes. We compute the logarithm of *Duration* to estimate an exponential model, ensuring that our estimates are not biased by experiences prior to observation period (Lapré and Tsikriktsis 2006).[10]

OP denotes a vector comprising operations performed on the same day before the focal operation. We consider three measures of team familiarity. First, we distinguish past operations based on the number of focal team members besides the surgeon. Our second measure is defined analogously, but groups operations performed by two or more focal team members besides the surgeon in the same category. Finally, we use the number of operations performed by any pair comprising the surgeon and any other member of the

[10]We obtain similar results with unlogged duration.

focal team. We note that the number of operations performed by the surgeon and one other member of the focal team is not equivalent to the number of operations conducted by any pair comprising the surgeon and any other member of the focal team.[11] A main objective of this paper is to compare the impact of these measures on team performance.

Besides team familiarity, **OP** accounts for experience accumulated individually by each of the team member on the day of the focal procedure. The surgeon's individual experience is captured by fixed effects indicating how many operations the surgeon performed prior to the focal procedure. For each team member besides the surgeon, we control for the number of operations performed without the focal surgeon to capture individual experience.[12]

X in (3.4) is a vector of characteristics of patient *i*. Demographic information includes 20-year age groups, gender, race, and marital status specified as fixed effects. To account for patient severity, we adjust for the primary ICD-9 procedure code of the operation performed by team *j* on patient *i*. In addition, we control for patient comorbidities by taking advantage of the Charlson comorbidity index (Charlson et al. 1987, Quan et al. 2011). The Charlson index is a widely used measure to adjust for the overall comorbidity severity of a patient.

Y denotes a vector of fixed effects for each member of team *j*, accounting for any skill variation that is constant during the observation period.[13] Finally, **T** denotes a vector comprising information about the timing of the operation of interest. It includes fixed effects for the hour of the day, the day of the week, and the month when the operation took place.[14] In all models, we cluster standard errors by surgeons to account for potentially correlated error terms.

[11]The first measure of team familiarity stems exclusively from operations performed by the surgeon and one — but no more — additional member of the focal team. By contrast, the pairwise measure of team familiarity also includes operations involving the surgeon and several other focal team members.

[12]We do not adjust for the number of operations performed by the surgeon without any of the focal team members. This variable is collinear once we control for fixed effects capturing the number operations performed by the surgeon on the day of the focal operation prior to the focal procedure and our measures of team familiarity.

[13]We specify residual fixed effects for members of the surgical team if these individuals appear fewer than 10 times in our final analysis sample.

[14]We specify a fixed effect for each month during the observation period. Thus, January 2008 is assigned to a different fixed effect than January 2009. Consequently, we do not use year fixed effects due to collinearity.

To explore whether team familiarity decays, we estimate (3.4) considering different time periods for the definition of **OP**. We consider operations performed on the same day, within 1–10 days, and within 11–20 days before the focal operation, respectively. We combine these measures in the same regression as follows:

$$\ln Duration_{ij} = \alpha + \beta \mathbf{OP}_j^{0} + \gamma \mathbf{OP}_j^{1-10} + \lambda \mathbf{OP}_j^{11-20} + \delta \mathbf{X}_i + \eta \mathbf{Y}_j + \theta \mathbf{T}_{ij} + \epsilon_{ij} \tag{3.5}$$

Here, the superscripts above **OP** denote the time period considered when computing the number of operations performed by one and two or more focal team members, respectively, besides the surgeon. As before, we control for fixed effects for the number of operations performed by the surgeon on the same day prior of the focal procedure. We also include the number of operations the surgeon performed with none of the focal team members within 1–10 and 11–20 days before the focal procedure and the number of operations performed by any other team member without the surgeon during any of the three time periods.

Comparing the β, γ, and λ coefficients shows to which extent team familiarity decays. We note that operations involving two or more focal team members are composed differently over time. That is, operations involving three or four of the team members besides the surgeon are far more frequent on the day of the focal procedure. These differences need to be taken into account when interpreting the coefficient estimates.

3.5 Results

3.5.1 Concentrated and Broad Team Familiarity

Table 3.3 reports how different types of operations affect procedure duration based on (3.4). In Column 1 we distinguish the surgeon's operations performed on the same day prior to the focal procedure based on the number of team members who are also involved in the focal operation. We find that an operation's impact on procedure duration steadily increases with the number of team members involved in both the past and the focal procedure. The reduction of procedure duration is especially pronounced between operations involving one

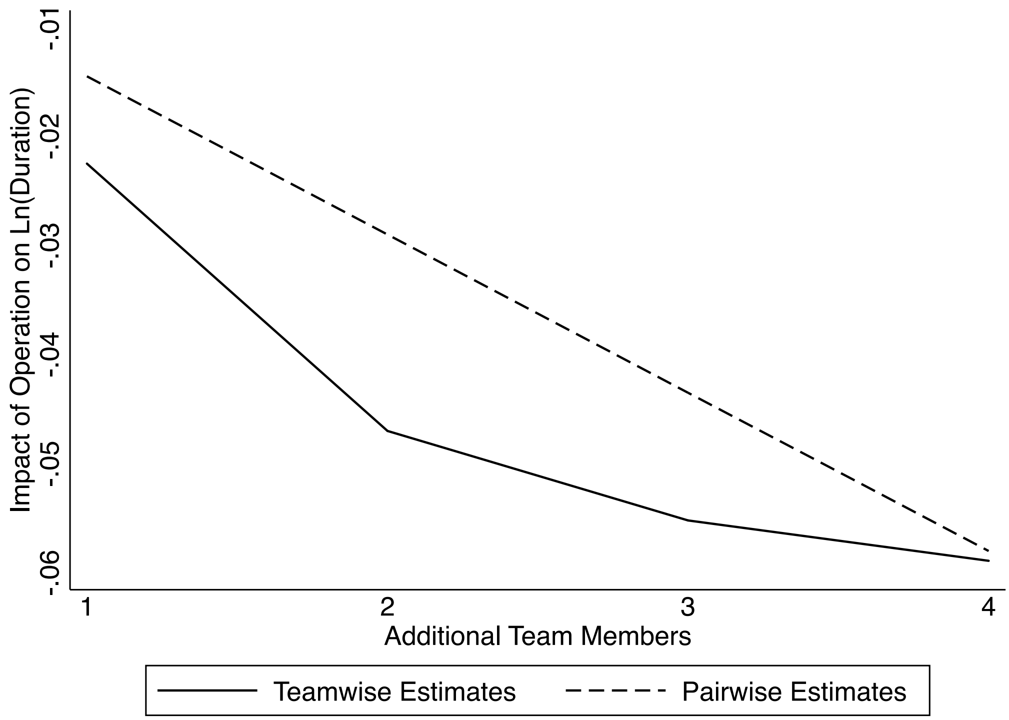
Table 3.3: *Team Familiarity and Procedure Duration*

	Ln(Duration)		
	(1)	(2)	(3)
Operations Surgeon & 1 Team Member	-0.0225***	-0.0219***	
	(0.0047)	(0.0047)	
Operations Surgeon & 2 Team Members	-0.0470***		
	(0.0065)		
Operations Surgeon & 3 Team Members	-0.0552***		
	(0.0047)		
Operations Surgeon & 4 Team Members	-0.0589***		
	(0.0055)		
Operations Surgeon & 2+ Team Members		-0.0533***	
		(0.0048)	
Operations Pair of Surgeon & Team Member			-0.0145***
			(0.0015)
Adjusted R2	0.7445	0.7445	0.7443
Observations	25180	25180	25180
Mean Ln(Duration)	4.7609	4.7609	4.7609

*Notes: This table reports how an operation performed previously on the same day impacts the duration of the focal procedure. We distinguish past operations based on the number of team members performing both the past and the focal procedure. Controls include patient 20-year age groups, gender, race, marital status, Charlson comorbidity index, primary ICD-9 procedure code as well as hour, weekday, and month-year fixed effects. For each team member, we include fixed effects and the number of operations not otherwise included. Models are estimated by OLS and standard errors are clustered by surgeon. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

and two team members, respectively, besides the surgeon. Surgeries involving three or four team members in addition to the surgeon do not reduce procedure duration by much more than surgeries involving the surgeon and two team members.

Column 2 combines operations involving two or more focal other team members besides the surgeon into a single operation category based on these findings. The resulting coefficient estimates are comparable to Column 1, suggesting that an operation involving one team member in addition to the surgeon — reflecting concentrated familiarity — reduces procedure duration by 2.2 percent. By contrast, broad familiarity — captured by operations involving two or more team members besides the surgeon — reduces procedure duration by 5.3 percent. Testing the hypothesis that these two coefficients are equal using a Wald test



Notes: This figure compares estimates of the impact of a prior operation on the day of the focal surgery on procedure duration. We distinguish past operations based on the number of team members performing both the past and the focal procedure. Teamwise estimates are reported in Column 1 of Table 3.3. Pairwise estimates are imputed from Column 3 of Table 3.3 by multiplying the coefficient estimate by the number of additional team members.

Figure 3.2: *Teamwise and Pairwise Estimates*

yields a F-statistic of 32.04 with a corresponding probability below one percent. Both concentrated and broad familiarity have a substantial impact on procedure duration, reducing it approximately by 3 and 7 minutes per operation, respectively.

Column 3 reports how the number of pairwise collaborations between the surgeon and any other focal team member on the same day prior to the focal operation affects procedure duration. This measure has been used in previous research to measure the familiarity of fluid teams (Reagans et al. 2005, Huckman et al. 2009). A pairwise operation has a significant effect on procedure duration, reducing it by 1.5 percent. Comparing Columns 1 and 3 enables us to compare pairwise and teamwise estimates as defined in Section 3.2.2. As highlighted in Figure 3.2, we find that pairwise estimates underestimate the "true" impact of team familiarity uncovered by teamwise estimates. The pairwise estimate documented in Column 3 is much smaller than the coefficient on operations of one team member besides the surgeon reported in Column 1. Similarly, the coefficients on operations involving the surgeon and more than one team member in Column 1 are always smaller than estimates obtained by multiplying the coefficient on pairwise operations by the corresponding number of team members, reflecting the number of additional pairwise collaborations.[15]

For robustness, we replicate the models presented in Columns 1 and 3 while weighing all team members equally rather than considering the surgeon as the central team member. We obtain coefficient estimates of -0.0235, -0.0447, -0.0564, and -0.0591 for operations involving two, three, four, or five of the focal team members, respectively. The coefficient estimate for the number of pairwise collaborations is -0.0047. All of these estimates are statistically significant at the one percent level. As before, the pairwise estimates do not capture the returns to broad and concentrated familiarity documented by the teamwise estimates.[16] For instance, the pairwise estimates underestimate the benefits to concentrated team familiarity by a factor of almost five. We conclude that the mathematical relationships proposed by (3.1) — at least in our setting — are not supported empirically.

3.5.2 Decay of Team Familiarity

In Table 3.4, we explore whether team familiarity decays.[17] Column 1 reports the impact of operations performed by the surgeon on the same day before the focal procedure. Both concentrated and broad team familiarity have a statistically significant impact on procedure

[15]Pairwise collaborations without the surgeon are irrelevant when the surgeon is considered the central team member. Thus, an operation involving two, three, or four team members increases the number of pairwise collaborations also by two, three, or four, respectively.

[16]According to the definition of (3.1), an operation involving two, three, four, and five team members leads to one, three, six, and ten additional pairwise collaborations. Multiplying the pairwise coefficient by these numbers results in estimates of -0.0047, -0.0141, -0.0282, and -0.0470, respectively. These numbers do not correspond to the estimates we obtain when allowing for concentrated and broad team familiarity.

[17]All specifications exclude observations for which operations performed 11–20 days before the focal procedure are undefined.

duration, reducing it by 2.1 and 5.2 percent for each corresponding operation, respectively. Column 2 considers operations performed within 1–10 days before the focal operation. Concentrated team familiarity reduces procedure duration by 0.1 percent. This effect is statistically insignificant. Meanwhile, broad familiarity reduces procedure duration significantly by 0.2 percent. Column 3 shows the estimates of an analogous specification for operations performed 11–20 days before the focal procedure. Here, concentrated familiarity appears to have a more substantial and significant impact than broad familiarity.

We combine operations performed during all time periods based on (3.5) in Column 4. We find that coefficients are broadly in line with the estimates documented in Columns 1 and 2, while the estimates analyzed individually in Column 3 appear smaller and less significant. Taken together, the estimates suggest that team familiarity developed on the day prior to the focal procedure is especially helpful in reducing procedure duration. Operations performed 1–10 days before the focal procedure are also beneficial, but at a far lower level.

Clearly, the considerable difference between the impact of operations performed on the same day and 1–10 days before the focal procedure could reflect an extraordinarily fast decay in team familiarity. Two alternative explanations may mediate the extent of decay we observe. First, operations involving three or four team members besides the surgeon are far more frequent on the day of the focal procedure. As these teams benefit slightly more from having collaborated in the past — suggested by the estimates in Column 1 of Table 3.3 — we would expect broad familiarity developed on the same day to be slightly more beneficial than broad familiarity developed 1–10 days before the focal procedure simply due to the composition of operations developing broad team familiarity. Second, there is some that returns to surgical team familiarity may be nonlinear (Xu et al. 2013). Thus, the higher number of procedures performed by surgeons and other team members within 1–10 days before the focal procedure may not translate to equivalent benefits in team familiarity.

Table 3.4: *Decay of Team Familiarity*

	Ln(Duration)			
	(1)	(2)	(3)	(4)
Operations Surgeon & 1 Team Member — 0 Days	-0.0213***			-0.0211***
	(0.0043)			(0.0041)
Operations Surgeon & 2+ Team Members — 0 Days	-0.0520***			-0.0514***
	(0.0049)			(0.0049)
Operations Surgeon & 1 Team Member — 1–10 Days		-0.0007		-0.0010*
		(0.0005)		(0.0005)
Operations Surgeon & 2+ Team Members — 1–10 Days		-0.0022**		-0.0024**
		(0.0011)		(0.0010)
Operations Surgeon & 1 Team Member — 11–20 Days			-0.0017**	-0.0011
			(0.0008)	(0.0007)
Operations Surgeon & 2+ Team Members — 11–20 Days			-0.0010	-0.0005
			(0.0019)	(0.0017)
Adjusted R2	0.7448	0.7409	0.7408	0.7449
Observations	24748	24748	24748	24748
Mean Ln(Duration)	4.7616	4.7616	4.7616	4.7616

*Notes: This table reports how a prior operations impacts the duration of the focal procedure. We distinguish past operations based on the number of team members performing both the past and the focal procedure. Controls include patient 20-year age groups, gender, race, marital status, Charlson comorbidity index, primary ICD-9 procedure code, operation room,as well as hour, weekday, and month-year fixed effects. For each team member, we include fixed effects and the number of operations not otherwise included. Models are estimated by OLS and standard errors are clustered by surgeon. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

Table 3.5: *Impact of Team Member Type on Team Familiarity*

	Ln(Duration)				
	(1)	(2)	(3)	(4)	(5)
Operations Surgeon & Assistant	-0.0318***				-0.0246***
	(0.0069)				(0.0067)
Operations Surgeon & Anesthesiologist		-0.0261***			-0.0156***
		(0.0040)			(0.0037)
Operations Surgeon & Scrub			-0.0273***		-0.0105**
			(0.0040)		(0.0052)
Operations Surgeon & Circulator				-0.0279***	-0.0092**
				(0.0042)	(0.0047)
Adjusted R2	0.7436	0.7433	0.7431	0.7432	0.7444
Observations	25180	25180	25180	25180	25180
Mean Ln(Duration)	4.7609	4.7609	4.7609	4.7609	4.7609

*Notes: This table reports how an operation performed previously on the same day impacts the duration of the focal procedure. We distinguish past operations based on the type of team member performing both the past and the focal procedure. Controls include patient 20-year age groups, gender, race, marital status, Charlson comorbidity index, primary ICD-9 procedure code as well as hour, weekday, and month-year fixed effects. For each team member, we include fixed effects and the number of operations not otherwise included. Models are estimated by OLS and standard errors are clustered by surgeon. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

3.5.3 Robustness

As discussed in Section 3.3, all of our analyses assume that the collaboration between any pair comprising the surgeon and any other team member affects team familiarity equally. We test the validity of this assumption in Table 3.5. In Column 1, we consider how procedure duration is affected by an operation that was performed within 10 days of the focal procedure and involved both the surgeon and the assisting surgeon. We obtain an significant estimate of -3.2 percent. We proceed similarly in Columns 2, 3, and 4, considering the impact of an operation involving the anesthesiologist, scrub nurse, and circulating nurse of the focal team, respectively, besides the surgeon performing the operation of interest. Procedure duration is reduced by 2.6–2.8 percent by these operations.

In Column 5, we include operations performed by all of these pairs in the same regression. Importantly, we cannot reject the hypothesis that an operation performed by any combination of the surgeon and any other member of the focal team reduces procedure duration equally. We note, however, that operations performed by the surgeon jointly with the assisting surgeon appear to build more team familiarity than other combinations.

We hypothesize that the impact of team familiarity on team performance may depend on physician experience and procedure complexity. Perhaps, experienced surgeons have learned to adapt faster to changing teams, resulting in less importance of team familiarity. On the other hand, surgeon experience may extend the degree to which surgeons benefit from team familiarity. For instance, surgeons may have to learn when they can trust their team. Complex procedures require better coordination among the team, making team familiarity very beneficial. Less complex procedures, however, offer a higher potential for standardization of processes within a team.

We use two measures to test whether our results are robust to these factors. The first proxy is the number of days a surgeon has practiced since her board certification. This measure capture a surgeon's overall experience with high precision, as the orthopedic surgeons we observe typically have been working full-time in teaching hospitals since the start of their career. In addition, we measure procedure complexity using the work relative

Table 3.6: *Impact of Physician and Procedure Characteristics on Team Familiarity*

	Ln(Duration)	
	(1)	(2)
Operations Surgeon & 1 Team Member	-0.0278***	-0.0241***
	(0.0093)	(0.0065)
Operations Surgeon & 2+ Team Members	-0.0508***	-0.0580***
	(0.0045)	(0.0053)
Operations Surgeon & 1 Team Member $*$ High Experience	0.0182	
	(0.0143)	
Operations Surgeon & 2+ Team Members $*$ High Experience	0.0064	
	(0.0095)	
Operations Surgeon & 1 Team Member $*$ High Complexity		0.0085
		(0.0095)
Operations Surgeon & 2+ Team Members $*$ High Complexity		0.0133*
		(0.0076)
Adjusted R2	0.7686	0.7510
Observations	21350	20532
Mean Ln(Duration)	4.7317	4.7428

*Notes: This table reports how the number of operations performed previously on the same day impacts the duration of the focal procedure. We distinguish past operations based on the type of team member performing both the past and the focal procedure. Experience is measured by the number of years since the physician's board certification. Complexity is measured by the work relative value unit used for Medicare reimbursement. Controls include patient 20-year age groups, gender, race, marital status, Charlson comorbidity index, primary ICD-9 procedure code as well as hour, weekday, and month-year fixed effects. For each team member, we include fixed effects and the number of operations not otherwise included. Models are estimated by OLS and standard errors are clustered by surgeon. ***, **, and * denote statistical significance at the one, five, and ten percent level, respectively.*

value unit — a reimbursement weight — defined by Medicare for each CPT codes. We link primary ICD-9 procedure codes to the most associated CPT code based on our data.

In Table 3.6, we interact concentrated and broad familiarity with high physician experience and procedure complexity. Physician experience and procedure complexity are defined to be high if they exceed the median value observed in our data. We distinguish operations based on the surgeon's experience in Column 1. Our findings suggest that both concentrated and broad team familiarity tends to matter more for inexperienced surgeons, although this effect is statistically insignificant. In Column 2, we explore the impact of procedure complexity on the relationship between team familiarity and procedure duration. We find that teams benefit more from both concentrated and broad team familiarity when they perform less complex procedures. Again, the results are statistically insignificant.

Finally, we confirm the robustness of our results to outliers by excluding operations of surgeons collaborating very frequently with other team members. To do so, we exclude operations if the surgeon's Herfindahl index with any type of team members is higher than the 95 percentile. This percentile is determined based on the distribution of Herfindahl indices with that team member type in our final analysis sample. Results remain unaffected by these exclusions. The coefficients on concentrated and broad team familiarity are -0.0218 and -0.0523, respectively, resembling the estimates of the analogous specification reported in Column 2 of Table 3.3 while being statistically significant at the one percent level.

3.6 Conclusions

We demonstrate that the familiarity of a surgical team — developed in past operations on the same day and 1–10 days before of the focal operation — reduces the duration of orthopedic surgeries. The impact of a past operation on procedure duration depends on the number team members associated with both that past operation and the focal operation. Broad familiarity shared among several of the focal team members affects procedure duration by more than twice as much as familiarity concentrated among two team members. We find that shared experience is especially beneficial on the day of the focal procedure and has no effect beyond 10 days before the focal operation.

The benefits to increasing a team's short-term familiarity are substantial. On the day of the focal procedure, a prior operation developing concentrated and broad familiarity reduces the expected procedure duration by about 3 and 7 minutes, respectively. Within 1–10 days before the focal procedure, an operation developing broad familiarity reduces procedure duration still by about 20 seconds on average. Teams frequently collaborate continuously on the day of the focal procedure, but only 10 percent of procedures performed within 1–10 days before the focal procedure develop broad familiarity.

Our analysis of the relative impact of concentrated and broad familiarity suggests that a team's performance increases nonlinearly with the volume of pairwise collaborations. We find that adding additional team members beyond two team members and the surgeon does not have substantial benefits. This result suggests that developing broad familiarity does note require having all team members working together. Meanwhile, the benefits of broad familiarity can be replicated by twice the amount of concentrated familiarity. Put differently, this paper provides some evidence for how team familiarity can be developed fairly efficiently in pieces. That is, having some members of the team work together is sufficient to achieve high levels of team familiarity.

In practice, this finding may have important implications. While it is faster to improve a team's performance by developing broad familiarity, building concentrated familiarity may be much more feasible. Across many settings, managers face a trade-off between developing team familiarity and attuning scheduling of various individuals. In some situations, focusing on the creation of concentrated rather than broad team familiarity may be optimal. Specifically, hospitals may find it much easier to align the schedules of two or three members of a surgical team rather than ensuring that a surgical team performs as many operations as possible in exactly the same composition.

Several limitations of our study deserve being mentioned. First, team composition in our setting is just quasi-random rather than perfectly random. To the extent that surgeons manage to obtain their preferred nurses and anesthesiologists, our results would be biased, reflecting team selection rather than team familiarity. We are not overly concerned about this issue, however, because of the specificities of our empirical setting and our focus on whole teams rather than specific team members. A second limitation of this analysis is the use of imprecise proxies, especially for procedure complexity. Despite finding inconclusive evidence, we do not rule out the possibility the team specificity of individual performance systematically varies with procedure complexity. Finally, more data in terms of years, surgeons, and medical disciplines would be very helpful in analyzing whether our results apply more broadly.

References

Almond, D., J. J. Doyle, Jr., A. E. Kowalski, and H. Williams (2010). Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns. *Quarterly Journal of Economics 125*(2), 591–634.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Anthony, D. L., M. B. Herndon, P. M. Gallagher, A. E. Barnato, J. P. W. Bynum, D. J. Gottlieb, E. S. Fisher, and J. S. Skinner (2009). How Much Do Patients' Preferences Contribute to Resource Use? *Health Affairs 28*(3), 864–873.

Arrow, H. and J. E. McGrath (1995). Membership Dynamics in Groups at Work: A Theoretical Framework. *Research in Organizational Behavior 17*, 373–373.

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review 53*(5), 941–973.

Baicker, K. and A. Chandra (2004). Medicare Spending, the Physician Workforce, and Beneficiaries' Quality of Care. *Health Affairs W4*, 184–197.

Bottle, A. and P. Aylin (2006). Mortality Associated with Delay in Operation after Hip Fracture: Observational Study. *BMJ 332*(7547), 947–951.

Brauer, C. A., M. Coca-Perraillon, D. M. Cutler, and A. B. Rosen (2009). Incidence and Mortality of Hip Fractures in the United States. *JAMA 302*(14), 1573–1579.

Chandra, A., D. Cutler, and Z. Song (2011). Who Ordered That? The Economics of Treatment Choices in Medical Care. In M. V. Pauly, T. G. McGuire, and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume 2, pp. 397–432. Amsterdam: Elsevier.

Chandra, A. and D. O. Staiger (2007). Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks. *The Journal of Political Economy 115*(1), 103–140.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie (1987). A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *Journal of Chronic Diseases 40*(5), 373–383.

Choudhry, N. K., G. M. Anderson, A. Laupacis, D. Ross-Degnan, S.-L. T. Normand, and S. B. Soumerai (2006). Impact of Adverse Events on Prescribing Warfarin in Patients with Atrial Fibrillation: Matched Pair Analysis. *BMJ 332*(7534), 141–145.

Clemens, J. and J. D. Gottlieb (2014). Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? *American Economic Review 104*(4), 1320–1349.

Cromwell, J. and J. B. Mitchell (1986). Physician-Induced Demand for Surgery. *Journal of Health Economics 5*(4), 293–313.

Cutler, D. M. (1995). The Incidence of Adverse Medical Outcomes Under Prospective Payment. *Econometrica 63*(1), 29–50.

Dafny, L. S. (2005). How Do Hospitals Respond to Price Changes? *American Economic Review 95*(5), 1525–1547.

Doyle, Jr., J. J. (2011). Returns to Local-Area Health Care Spending: Evidence from Health Shocks to Patients Far From Home. *American Economic Journal: Applied Economics 3*(3), 221–243.

Doyle, Jr., J. J., S. M. Ewer, and T. H. Wagner (2010). Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams. *Journal of Health Economics 29*(6), 866–882.

Dranove, D. and P. Wehner (1994). Physician-Induced Demand for Childbirths. *Journal of Health Economics 13*(1), 61–73.

Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly 44*(2), 350–383.

Edmondson, A. C., R. M. Bohmer, and G. P. Pisano (2001). Disrupted Routines: Team Learning and New Technology Implementation in Hospitals. *Administrative Science Quarterly 46*(4), 685–716.

Edmondson, A. C. and I. M. Nembhard (2009). Product Development and Learning in Project Teams: The Challenges Are the Benefits. *Journal of Product Innovation Management 26*(2), 123–138.

Edmondson, A. C., A. B. Winslow, R. M. Bohmer, and G. P. Pisano (2003). Learning How and Learning What: Effects of Tacit and Codified Knowledge on Performance Improvement Following Technology Adoption. *Decision Sciences 34*(2), 197–224.

Ellis, R. P. and T. G. McGuire (1986). Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply. *Journal of Health Economics 5*(2), 129–151.

Enthoven, A. C. (1980). *Health Plan: The Only Practical Solution to the Soaring Cost of Medical Care*. Reading, MA: Addison-Wesley.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and É. L. Pinder (2003). The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care. *Annals of Internal Medicine 138*(4), 288–298.

Fox, H., J. Pooler, D. Prothero, and G. Bannister (1994). Factors Affecting the Outcome after Proximal Femoral Fractures. *Injury 25*(5), 297–300.

Fuchs, V. R. (1978). The Supply of Surgeons and the Demand for Operations. *Journal of Human Resources 13*, 35–56.

Fuchs, V. R. (2004). More Variation in Use of Care, More Flat-Of-The-Curve Medicine. *Health Affairs VAR*, 104–107.

Glover, J. A. (1938). The Incidence of Tonsillectomy in School Children. *Proceedings of the Royal Society of Medicine 31*(10), 1219–1236.

Gruber, J. and M. Owings (1996). Physician Financial Incentives and Cesarean Section Delivery. *RAND Journal of Economics 27*(1), 99–123.

Holmberg, S., R. Kalén, and K.-G. Thorngren (1987). Treatment and Outcome of Femoral Neck Fractures: An Analysis of 2418 Patients Admitted from Their Own Homes. *Clinical Orthopaedics and Related Research 218*, 42–52.

Huckman, R. S. and G. P. Pisano (2006). The Firm Specificity of Individual Performance: Evidence from Cardiac Surgery. *Management Science 52*(4), 473–488.

Huckman, R. S. and B. R. Staats (2011). Fluid Tasks and Fluid Teams: The Impact of Diversity in Experience and Team Familiarity on Team Performance. *Manufacturing & Service Operations Management 13*(3), 310–328.

Huckman, R. S., B. R. Staats, and D. M. Upton (2009). Team Familiarity, Role Experience, and Performance: Evidence from Indian Software Services. *Management Science 55*(1), 85–100.

Katz, R. (1982). The Effects of Group Longevity on Project Communication and Performance. *Administrative Science Quarterly 27*(1), 81–104.

Kessler, S., S. Kinkel, W. Kafer, W. Puhl, and T. Schochat (2003). Influence of Operation Duration on Perioperative Morbidity in Revision Total Hip Arthroplasty. *Acta Orthopaedica Belgica 69*(4), 328–333.

Khan, S. K., S. Kalra, A. Khanna, M. M. Thiruvengada, and M. J. Parker (2009). Timing of Surgery for Hip Fractures: A Systematic Review of 52 Published Studies Involving 291,413 Patients. *Injury 40*(7), 692–697.

Lapré, M. A. and N. Tsikriktsis (2006). Organizational Learning Curves for Customer Dissatisfaction: Heterogeneity across Airlines. *Management Science 52*(3), 352–366.

Leong, G., J. Wilson, and A. Charlett (2006). Duration of Operation as a Risk Factor for Surgical Site Infection: Comparison of English and US Data. *Journal of Hospital Infection 63*(3), 255–262.

McClellan, M. (2011). Reforming Payments to Healthcare Providers: The Key to Slowing Healthcare Cost Growth while Improving Quality? *Journal of Economic Perspectives 25*(2), 69–92.

McClellan, M., B. J. McNeil, and J. P. Newhouse (1994). Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis using Instrumental Variables. *JAMA 272*(11), 859–866.

Moran, C. G., R. T. Wenn, M. Sikand, and A. M. Taylor (2005). Early Mortality after Hip Fracture: Is Delay before Surgery Important? *Journal of Bone & Joint Surgery, American 87*(3), 483–489.

O'Hare, A. M., R. A. Rodriguez, S. M. Hailpern, E. B. Larson, and M. K. Tamura (2010). Regional Variation in Health Care Intensity and Treatment Practices for End-Stage Renal Disease in Older Adults. *JAMA 304*(2), 180–186.

Okike, K., O. C. Lee, H. Makanji, M. B. Harris, and M. S. Vrahas (2013). Factors Associated with the Decision for Operative versus Non-Operative Treatment of Displaced Proximal Humerus Fractures in the Elderly. *Injury 44*(4), 448–455.

Ong, K. L., E. Lau, M. Manley, and S. M. Kurtz (2008). Effect of Procedure Duration on Total Hip Arthroplasty and Total Knee Arthroplasty Survivorship in the United States Medicare Population. *Journal of Arthroplasty 23*(6), 127–132.

Pisano, G. P., R. M. Bohmer, and A. C. Edmondson (2001). Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery. *Management Science 47*(6), 752–768.

Quan, H., B. Li, C. M. Couris, K. Fushimi, P. Graham, P. Hider, J.-M. Januel, and V. Sundararajan (2011). Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data from 6 countries. *American Journal of Epidemiology 173*(6), 676–682.

Reagans, R., L. Argote, and D. Brooks (2005). Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together. *Management Science 51*(6), 869–881.

Rogers, F. B., S. R. Shackford, and M. S. Keller (1995). Early Fixation Reduces Morbidity and Mortality in Elderly Patients with Hip Fractures from Low-Impact Falls. *Journal of Trauma-Injury, Infection, and Critical Care 39*(2), 261–265.

Schilling, M. A., P. Vidal, R. E. Ployhart, and A. Marangoni (2003). Learning by Doing Something Else: Variation, Relatedness, and the Learning Curve. *Management Science 49*(1), 39–56.

Shiga, T., Z. Wajima, and Y. Ohe (2008). Is Operative Delay Associated with Increased Mortality of Hip Fracture Patients? Systematic Review, Meta-Analysis, and Meta-Regression. *Canadian Journal of Anesthesia 55*(3), 146–154.

Sinclair, G., S. Klepper, and W. Cohen (2000). What's Experience Got to Do with It? Sources of Cost Reduction in a Large Specialty Chemicals Producer. *Management Science 46*(1), 28–45.

Sommers, B. D., C. J. Beard, A. V. D'Amico, I. Kaplan, J. P. Richie, and R. J. Zeckhauser (2008). Predictors of Patient Preferences and Treatment Choices for Localized Prostate Cancer. *Cancer 113*(8), 2058–2067.

Song, Y., J. Skinner, J. Bynum, J. Sutherland, J. E. Wennberg, and E. S. Fisher (2010). Regional Variations in Diagnostic Practices. *New England Journal of Medicine 363*(1), 45–53.

Staiger, D. and J. H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica 65*(3), 557–586.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics 20*(4), 518–529.

Welch, H. G., S. M. Sharp, D. J. Gottlieb, J. S. Skinner, and J. E. Wennberg (2011). Geographic Variation in Diagnosis Frequency and Risk of Death among Medicare Beneficiaries. *JAMA 305*(11), 1113–1118.

Weller, I., E. Wai, S. Jaglal, and H. Kreder (2005). The Effect of Hospital Type and Surgical Delay on Mortality after Surgery for Hip Fracture. *Journal of Bone & Joint Surgery, British 87*(3), 361–366.

Wennberg, J. and A. Gittelsohn (1973). Small Area Variations in Health Care Delivery. *Science 182*(4117), 1102–1108.

Williams, Gerald R., J., C. A. Rockwood, Jr., L. U. Bigliani, J. P. Iannotti, and W. Stanwood (2004). Rotator Cuff Tears: Why Do We Repair Them? *Journal of Bone & Joint Surgery, American 86*(12), 2764–2776.

Wright, P. G. (1928). *Tariff on Animal and Vegetable Oils*. New York, NY: MacMillan.

Xu, R., M. J. Carty, D. P. Orgill, S. R. Lipsitz, and A. Duclos (2013). The Teaming Curve: A Longitudinal Study of the Influence of Surgical Team Familiarity on Operative Time. *Annals of Surgery 258*(6), 953–957.