NORTHWESTERN UNIVERSITY

Empirical Studies on the Organization of Health Care

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Managerial Economics and Strategy

By

Subramaniam Ramanarayanan

EVANSTON, ILLINOIS

December 2007

ABSTRACT

Empirical Studies on the Organization of Health Care

Subramaniam Ramanarayanan

This thesis contains four essays on the organization (both internal and external) of health care in the US. The first essay examines a mechanism through which individual workers acquire (or maintain) competence, namely that of experience. Specifically, I analyze whether cardiac surgeons who perform more procedures experience an improvement in performance. As my identification strategy, I consider exogenous shocks to the procedure volume of CABG surgeons in Florida caused by the exit of other surgeons from the same hospital. Using this instrument, I find evidence indicating a strong learning-by-doing effect for cardiac surgeons: an additional procedure a year leads to a reduction in the probability of patient mortality by .05 percentage points. Further, I find this improvement in surgeon performance to be completely transferable across different hospital settings, and find evidence of some economies of scope among the different surgical procedures performed by a cardiac surgeon.

The second and third essays also deal with physician behavior – focusing on their medical practice style, which is defined as the propensity of the physician to prescribe a particular treatment to a given patient. In the second essay (joint with David Dranove and Hayagreeva Rao), we propose several new statistical methods for assessing the presence and sources of small area variations in physician practice. In the third essay (joint with David Dranove), we focus on "credence good" markets and attempt to answer the question: what prevents agents from always recommending more costly services? We examine whether the market punishes agents who are deemed to be too "aggressive". Focusing on Ob/Gyn physicians, we find that maternity patients prefer not to visit physicians who overprescribe cesarean sections, i.e. who are more aggressive.

The final essay (joint with Leemore Dafny and Mark Duggan) looks at the extent to which rising premiums in the private health insurance industry can be explained by the increasing consolidation among insurance carriers. We make use of a large, private database of insurance contracts in order to establish this link. By looking at within-market variation, we find that a one standard-deviation change in concentration leads to a premium increase of 2.5-4%.

Acknowledgements

This dissertation is the product of a fantastic educational experience at Northwestern University. Various people have contributed towards its completion at different stages, either directly or indirectly, and any attempt to thank all of them is bound to fall short. To begin, I would like to thank my instructors at the Indian Institute of Management, Calcutta – Prof B.B. Chakrabarti and Prof Anup Sinha in particular - for encouraging me to attend graduate school.

I owe a great deal to my dissertation committee. David Dranove has been a strong influence on the way I approach empirical research and on my choice of research topics. He has been a constant source of encouragement and has provided me a lot of research and teaching opportunities for which I am especially grateful. Leemore Dafny is a terrific role model and a big source of inspiration. She made sure I was on track, constantly pushed the boundaries of my research and was always present to respond to my queries, no matter how insignificant they were. Her advice and support have gone a long way towards making me the academic researcher that I am today. David Besanko was a great source of encouragement and of great help during the job market. Huggy (Hayagreeva) Rao's sunny disposition, intellectual curiosity and breadth of knowledge still leave me awestruck at times. I thank him for all the candid advice and the tips and for his continued interest in my work even after moving to Stanford.

Doctoral research is often a lonely endeavor – thanks to my friends and classmates at Northwestern, it rarely felt that way. Ron and Sourav, with whom I shared an office for five years, will be sorely missed. I would also like to thank Caroline, Ewa, Min, Mian, Seongwuk, Susan and Tapas for the innumerable cups of coffee and for all the interesting conversation. Thanks to Urmi and Gayle for introducing me to the laidback lifestyle of graduate school. Thanks are also due to a few institutions in the Magnificent City of Chicago, which helped the five years fly by – the Radio Jockeys on the Drive (97.1 FM) for their wonderful taste in music, the baristas at the Brothers K CoffeeHouse for their cappuccinos and spinach pies, and the fine folks at Bar Louie, Evanston.

I am blessed to have a family that strongly believes in me and has always given me their unconditional love and support. My parents – Ram and Satya – taught me everything I know about the values of hard work, integrity and dedication. I am what I am today because of them. I was fortunate to be located close to part of my family - Karthik, Swetha and Anoushka - they made me feel at home all these years.

Finally, I dedicate this dissertation to Neelam, my fiancée, who has been my biggest source of inspiration. Her enthusiasm for research and commitment towards her work has always inspired me to reach further. She has been a constant source of strength, laughter and good advice and now probably knows far more about a PhD in Economics than she ever wished she would. I could not have completed this effort without her.

Contents

List of Tables

1 Introduction

This dissertation contains four essays that look at different aspects of the way health care is organized in the US. The first three essays (Chapters 2-4) examine physician behavior while the final essay studies the industrial organization of the private health insurance market. The rest of this introduction provides an overview of each of the four chapters, touching on motivations, methodologies and results.

1.1 The Impact of Experience on Performance of Individual Workers

The effects of organizational experience on quality and costs have been studied in multiple settings. The realization of productivity gains with increased experience is termed learning-by-doing and the presence of a learning curve has been well documented in manufacturing and service firms. However, the extent of our knowledge about why learning occurs is still limited. For example, learning could result from the acquisition of human capital by individual employees or from increased organizational experience that helps standardize production processes. In this paper, I analyze one of the determinants of worker competence in an organizational setting: task experience. In particular, I estimate the extent to which a surgeon's recent procedure volume (measured as the number of CABG surgeries performed the previous year) affects performance (measured by patient outcomes). In addition, the setting of this study enables me to test for the degree of specificity (to a firm and to a task) of human capital that is acquired by individuals through learning-by-doing.

Focusing on individual experience provides interesting insights into the learning process of firms. The results from this study can also help shape a firm's strategy on hiring and retention of skilled workers and on allocation of workload across workers. It is also an important question to examine in the health care setting given that provider experience is often taken as a proxy for quality. However, there is surprisingly little convincing empirical evidence that establishes the presence of learning-by-doing in this setting. Estimates from prior studies are confounded by a fundamental problem of identification: Does experience result in learning or does quality reflect unobserved skills resulting in greater demand, and thus greater experience?

I use an instrumental variables approach in this study to establish a causal relationship between surgeon procedure volume and patient outcomes. I propose the use of surgeon exit as an exogenous identifier, where the term "exit" connotes any instance where a surgeon stops performing CABG surgeries in Florida. I use data on surgeon characteristics to restrict the set of exitors to surgeons aged 55 and over. By looking at the change in surgeon quality resulting from this shock, I am able to disentangle the learning-by-doing effect from other potential explanations underlying the volume-outcome relationship.

I find evidence in support of a strong learning-by-doing effect for CABG surgeons: performing a single additional procedure in the prior year yields a reduction in the probability of an adverse patient outcome of .051 percentage points, which translates to a 1.2% drop (relative to the average mortality rate for CABG of 3.82%). This result is fairly robust to alternate definitions of exit, alternate empirical specifications and to different methods of allocating the procedure volume of exitors among non-exiting surgeons. Also, surgeons performing a large volume of procedures benefit less from additional experience when compared to surgeons with low procedure volume.

1.2 Understanding the Source of Medical Practice Variations

Numerous studies have shown that the type of treatment a patient receives may depend on where the patient lives, and not just on what condition the patient has. Many scholars have argued that such small area variations (SAVs) in practice styles[1] imply market failure and lead to welfare losses. A number of traditional explanations for SAVs such as differences in patient preferences, income, underlying health status of physicians, physician density, access to medical care, availability of substitutes, or noise stemming from sampling have been disconfirmed in the literature.

In this study, we then assess whether SAVs are artifacts of individual-level variations among physicians, or artifacts of hospital-level variations, and address this issue using a simple but novel method that can be easily generalized to other procedures/regions. We specifically ask "Is the measured CoV when physicians and their patients are grouped by predetermined geographic region larger than the CoV obtained when physicians and hospitals are randomly assigned to "pseudo" regions that are equal in number and makeup to the actual regions?" We find that the median CoV in the actual data is roughly comparable to the median CoV's obtained through the random assignment of hospitals to pseudo-counties. This suggests that when it comes to c-sections in Florida, there do not appear to be any meaningful *regional* practice style effects; all variations in practice style are limited to variations across *hospitals*. We then ask why are physician styles correlated by hospital. We are able to conclude in favor of a substantial matching effect, and are able to rule out a learning effect, i.e. physician styles seem to be pretty stable over time

[1] We use the term practice style to connote the propensity of a physician to treat a given patient in a particular way. For example, some physicians may have a propensity to deliver babies by caesarian section whereas others may favor vaginal delivery, all else equal.

1.3 Why don't agents always recommend more costly services?

In "credence good" markets (e.g. plumbers, auto repair, doctors), customers have limited information on which to judge the merits of the recommendation offered by an expert and may, as a result, consent to excessively costly or unnecessary services. What prevents agents from always recommending more costly services? There are some obvious constraints such as the face validity of the recommendation (such as if a patient admits to mild acid reflux and the physician recommends a heart transplant) and the potential for litigation (e.g., malpractice after unnecessary surgery is botched). Within health economics, McGuire and Pauly (1991) suggest that physicians are constrained by ethical considerations. Also writing about demand inducement in health economics, Dranove (1988) suggests patients may avoid physicians who have a reputation for doing too many high cost procedures. In other words, simple market forces may limit the ability of sellers to exploit their informational advantage.

In this paper, we look at whether the market punishes overzealous sellers. Focusing on the market for deliveries, we estimate a model of consumer choice, where one of the factors weighing on patient's choice of provider is that provider's practice style. We find that maternity patients prefer not to visit physicians with aggressive styles (i.e. those who over prescribe cesarean sections), *ceteris paribus*. The effect is most pronounced for high income patients and HMO patients, two segments of the market that might be very attractive to some obstetricians.

In ongoing work, we are trying to resolve an identification problem with our current analyses: while we interpret our findings as reflecting consumer distaste for aggressive physicians, an alternative explanation could be that physicians are inducing demand for cesarean sections in the face of falling market share. We are currently trying out various solutions, including an instrumental variables approach, to resolve this problem.

1.4 Does Insurer Consolidation Lead to Higher Premiums?

In a recent paper, Robinson (2004) documents the existence of two, possibly related, trends in the private health insurance industry: the increase in local market concentration, and the growth in insurer profits and premiums over the last few years. No empirical research has shown a causal link between these trends. Using a large, private database of insurance contracts representing 10+ million covered lives, we establish that link. Our analysis uses panel data on local insurance markets from 1998-2005. We examine the effects of changes in local market concentration on changes in insurance premiums for fully-insured HMO plans. The former is instrumented using the number of insurance carriers in a market.

The estimates from market-level OLS as well as IV regressions provide strong support for the hypothesis that increased market concentration of insurers leads to higher prices. Depending on the specification and the instrument used, the IV results imply a 2.5-4% increase in health insurance premiums for a one standard deviation increase in the measure of market concentration. We are currently working on extending the set of instruments by using mergers between national insurance carriers as a proxy for concentration, the underlying logic being that consolidation of these carriers is less likely to be correlated with trends in insurance premiums in any particular market.

2 Does Practice Make Perfect: An Empirical Analysis of Learning-by-Doing in Cardiac Surgery

2.1 Introduction

The effects of organizational experience on quality and costs have been studied in multiple settings. The realization of productivity gains with increased experience is termed learning-by-doing and the presence of a learning curve has been well documented in manufacturing and service firms. However, the extent of our knowledge about why learning occurs is still limited. For example, learning could result from the acquisition of human capital by individual employees or from increased organizational experience that helps standardize production processes. In this paper, I analyze one of the determinants of worker competence in an organizational setting: task experience. I use the term experience here to connote frequency of performing a certain task, as opposed to number of years spent on the job.

In industries employing skilled labor, learning is thought to result mainly from workers becoming more efficient at the tasks they perform through multiple repetitions. I study learning-by-doing at the level of the individual worker in one such setting, specifically that of cardiac surgeons performing Coronary Artery Bypass Graft (CABG) surgeries in hospitals. In particular, I estimate the extent to which a surgeon's recent procedure volume (measured as the number of CABG surgeries performed the previous year) affects performance (measured by patient outcomes). In addition, the setting of this study enables me to test for the degree of specificity (to a firm and to a task) of human capital that is acquired by individuals through learning-by-doing.

Focusing on individual experience provides interesting insights into the learning process of firms. Recent research in the field of organizational learning has focused on the differences in learning rates across organizations (e.g. Pisano et al (2001); Reagans et al (2005)), and proficiency of individual workers has been identified as one of the factors driving this variation. The results from this study can also help shape a firm's strategy on hiring and retention of skilled workers and on allocation of workload across workers. In the health care setting, provider experience has long been used as a proxy for quality by patients, health care providers, surgical accreditation boards and insurers. For instance, the American College of Surgeons states in its guidelines for Coronary Artery Bypass Grafts (CABG) that in order to maintain competence, a surgeon should perform a minimum of 200 procedures per year as the primary operator. Procedure volume also plays a prominent role in hospital marketing brochures and websites, suggesting that it is a metric widely used within the industry to signal quality.

Given the significance attached to provider experience in health care, there is surprisingly little convincing empirical evidence that establishes the presence of learning-by-doing in this setting. Estimates from prior studies are confounded by a fundamental problem of identification: Does experience result in learning or does quality reflect unobserved skills resulting in greater demand, and thus greater experience? In other words, do surgeons improve with experience or are high quality surgeons more experienced because they attract more patients? Existing studies in the healthcare literature assume the presence of a correlation between surgeon experience and quality as being indicative of a learning effect. Failure to account for reverse causality in these papers leads to biased estimates of learning-by-doing.

I use an instrumental variables approach in this study to establish a causal relationship between surgeon procedure volume and patient outcomes. I propose the use of surgeon exit as an exogenous identifier, where the term "exit" connotes any instance where a surgeon stops performing CABG surgeries in Florida. I use data on surgeon characteristics to restrict the set of exitors to surgeons aged 55 and over. Doing so mitigates concerns about possible endogeneity of exit. The rationale behind the instrument is as follows: once exit occurs, patients who would have been treated by the exiting surgeon are now allocated among the remaining (non-exiting) surgeons at the hospital, thereby providing a positive shock to their procedure volumes. By looking at the change in surgeon quality resulting from this shock, I am able to disentangle the learning-by-doing effect from other potential explanations underlying the volume-outcome relationship. I also provide empirical evidence in support of the relevance and exogeneity of the instrument, i.e. I show that the volume shock resulting from exit acts as a good predictor of expected change in (staying) surgeon procedure volumes and is exogenous to changes in surgeon quality.

I make use of a patient-level dataset from the Agency of Health Care Administration (AHCA) in Florida that identifies the unique license number of the operating surgeon. By using the volume of exiting surgeons as an instrument for surgeon procedure volume, I find evidence in support of a strong learning-by-doing effect for CABG surgeons: performing a single additional procedure in the prior year yields a reduction in the probability of an adverse patient outcome of .051 percentage points, which translates to a 1.2% drop (relative to the average mortality rate for CABG of 3.82%). This result is fairly robust to alternate definitions of exit, alternate empirical specifications and to different methods of allocating the procedure volume of exitors among non-exiting surgeons. Also, surgeons performing a large volume of procedures benefit less from additional experience when compared to surgeons with low procedure volume.

I extend this methodology to examine the degree of specificity (to a firm and to a task) of human capital acquired by surgeons as a result of learning-by-doing. In order to test for firm-specificity, I exploit the fact that cardiac surgeons are affiliated with (and perform procedures at) multiple hospitals at the same point in time. This allows me to use hospital-specific measures of surgeon experience[2] and enables me to test whether surgeon performance is affected more by experience at the current hospital than by experience at other hospitals. In order to resolve problems of endogeneity, I instrument for hospital-specific experience using a simple modification of the instrument described above. The estimates from this specification indicate that the improvement in surgeon performance is completely transferable across different hospital settings. Since cardiac surgeons perform procedures other than CABG surgeries, I am also able to test whether experience with non-CABG procedures benefits surgeon performance in CABG by using an estimation strategy similar to the one used to test for firm-specificity. The results indicate that greater experience performing non-CABG procedures does benefit outcomes of CABG patients, but this effect is smaller than the effect of experience performing CABG procedures, i.e. there is some degree of task-specificity to the human capital of surgeons gained from procedure experience.

Taken together, these findings have important implications for managers within the hospital industry, specifically with respect to how to organize the firm so as to better leverage the skills of its professional workers. While allocating procedure volume within a team, it is important for managers to bear in mind that the marginal benefit of additional experience is greater for low volume surgeons when compared to high volume surgeons. Hospitals should ensure that low-volume surgeons get to develop mentoring relationships with more experienced surgeons. The findings on firm- and task-specificity have implications for hiring and retention policies and optimal job design, respectively.

---

[2] Huckman and Pisano (2006) were the first to test for firm-specificity of experience in this setting. In Section 2.3, I outline the main differences between their approach and mine.

This study contributes to existing empirical studies on learning-by-doing, especially those with a focus on health care (e.g. Hughes et al (1987); Hannan et al (1991); Gowrisankaran et al (2006); Ho (2002)). Most existing studies in this literature fail to adequately account for alternate explanations underlying the observed correlation between experience and outcomes. The primary contribution of this paper is that it is able to establish a causal link between provider experience and patient outcomes by using an instrumental variables technique. Also, the lengthy study period (1998-2003) helps me overcome some of the data limitations experienced by prior researchers. The study also complements recent theoretical and empirical research on firm-specific and task-specific aspects of human capital (e.g. Huckman and Pisano (2006); Gibbons and Waldman (2004)).

The rest of this paper is organized as follows. The next section provides a review of three streams of literature related to this study, and outlines some of the shortcomings of existing empirical research on learning-by-doing in health care. Section 2.3 describes the data and provides some institutional background. Section 2.4 outlines the empirical approach followed and describes the instrument in detail. Section 2.5 presents the main empirical specifications. The final two sections contain a discussion of results and some concluding remarks.

2.2 Background and Related Research

The present study draws from multiple streams of related research – studies of learning-by-doing in the industrial organization and strategy literatures, human capital theory and volume-outcome studies in the fields of medicine and health economics. I use insights from all these literatures in my model, and aim to make a contribution to each of them.

2.2.1 The Benefits of Experience

Starting with Wright's (1936) seminal analysis of airframe production, there has been a significant body of empirical and theoretical research aimed at documenting the association between cumulative experience (typically measured by cumulative production volume) and performance improvement across different industries. Some recent settings in which learning curves have been studied include aircraft production (Benkard (2000)), shipbuilding (Argote, Beckman and Epple (1990), Thompson (2001, 2006)), semiconductors (Hatch and Mowery (1998)) and pizza franchises (Darr, Argote and Epple (1995)).[3] These firm-level studies examine the impact of learning-by-doing using unit costs as a measure of performance. I exploit the availability of detailed micro-level data on patients undergoing CABG surgery to study the process of learning at the level of the individual worker. Also, unlike the studies mentioned above, I characterize performance in terms of patient outcomes (or surgeon quality), which, unlike costs, are measured with greater precision.

The effect of experience on individual productivity has been investigated to a lesser extent in these literatures.[4] In the setting I examine, there are a number of ways in which individual experience could matter in determining performance. The skill of the operating CABG surgeon is critical in ensuring a successful procedure. According to the American College of Cardiology[5], experience has a strong influence on a surgeon's cognitive knowledge base and technical skills, both of which determine competence. Experience helps surgeons in selecting the appropriate treatment strategy for patients, and helps them identify and treat complications at an early stage. Experience also has a positive impact on manual dexterity and helps surgeons maintain proper surgical technique. Finally, because of the rare occurrence of adverse outcomes, surgeon competence requires specific training and ongoing experience in managing them so as to be prepared to react optimally when they occur.

---

[3] Some of these studies (e.g. Benkard (2000)) also explicitly model the effects of organizational "forgetting".

[4] The literature in psychology has numerous studies on this effect. These studies typically find that individuals complete a task in lesser time and with greater accuracy, the more experience they have with the task. They also find support for diminishing returns of experience. Examples include Newell and Rosenbloom (1981), Delaney, Reder, Staszewski and Ritter (1998).

[5] Refer to Hirshfeld et al (1998) for the complete statement of clinical competence released by the American College of Cardiology.

2.2.2 The Theory of Human Capital

The second stream of research relevant to this study examines the nature of accumulation of human capital by individuals. This literature (Becker (1962), Killingsworth (1982)) presents two ways in which individuals accumulate human capital: training (education or on-the-job) and learning-by-doing. In the context studied here, a surgeon may increase her human capital by investing in (residency) training. She may also accumulate human capital as a result of the skills she gains from performing more procedures.

One aspect of human capital emphasized in this literature from very early on (Becker (1962), Jovanovic (1979)) is the importance of assessing the degree of specificity. Becker (1962) focuses on the dichotomy between general human capital (augments productivity across all firms) and firm-specific human capital (augments productivity in current firm, but not elsewhere). Recent research (Gibbons and Waldman (2004)) has focused on a third category: task-specific human capital, which is defined as being specific to the tasks being performed, as opposed to being specific to a firm. The data in this study allow me to estimate whether human capital acquired by individuals through learning-by-doing is specific to (a) the firms they work in and (b) the tasks they perform. In doing so, this study adds to the empirical evidence establishing learning-by-doing as a source of general or specific human capital. Huckman and Pisano (2006), who study the firm-specificity of performance of cardiac surgeons, is a recent example of one such study. I discuss their approach in detail in the next subsection.

For a cardiac surgeon, general human capital would result from investment in medical school, residency and post-residency training. Another determinant of the level of human capital could be procedure experience, which may help the surgeon develop new skills or maintain old ones. Studies in the human capital literature typically use earnings or wages as their measure of productivity. Since

I do not have access to earnings data for surgeons, I use surgeon performance to test for specificity of human capital.

Why may one expect accumulated human capital to be (completely or partially) specific to a firm (hospital) in this setting? Huckman and Pisano (2006) offer some potential explanations, the primary one being that a surgeon tends to develop a degree of familiarity with the rest of the surgical team and other assets at a hospital and this aspect prevents the benefits of experience from being portable across hospitals. On the other hand, if familiarity with assets were not as important as, say, learning how to treat complications, one can imagine a situation where experience adds to both general and firm-specific aspects of human capital, but the general component is much larger.

It is also plausible to think of a scenario where part of the human capital accumulated by surgeons on the job is specific to the tasks being performed, as opposed to being specific to the firm. In this context, I use the term "task" to connote a specific procedure. A cardiac surgeon may perform multiple tasks within a firm: she may perform CABG surgeries and she may also perform other cardiac procedures (e.g. repair of heart valves, heart transplants etc.). In such a situation, some part of the human capital may go unused when the surgeon switches tasks (either within the firm or across firms). Whether surgeon learning transfers across tasks (procedures) depends on the extent to which the skills (of the surgeon or the surgical team) needed to perform these procedures overlap.

Sections 2.5 and 2.6 contain the model and results for the tests of these hypotheses.

2.2.3 Volume Outcome Studies in Healthcare

The final stream of research relevant to this study encompasses papers in the medical and health economics literatures that document a correlation between procedure volume and outcomes at the firm (hospital) and individual (surgeon) levels. While it is difficult to compare findings across studies because of differences in data, disease categories studied and methodology, the general consensus among researchers seems to be that there is a positive relationship between procedure volume (measured at the level of the hospital or the individual surgeon) and patient outcomes.[6]

However, few studies attempt to translate this correlation into a well established causal relationship between procedure volume and outcome. The literature offers two competing hypotheses, with contrasting causal and policy implications that could underlie this observed relationship. The first hypothesis is the "learning-by-doing" or the "practice-makes-perfect" hypothesis. This hypothesis is built on the notion that increased experience results in more finely developed skills which in turn lead to better outcomes.

The second hypothesis is the "selective referral" hypothesis, which postulates that the observed relationship is due to a referral system that directs more patients to high quality providers. Even if patients do not have knowledge about exact mortality rates of a surgeon's former patients, one can think of a scenario where the best surgeons and the best hospitals develop a reputation as being of high quality, and hence attract more patients or referrals from specialists and primary care physicians.

The need to unambiguously distinguish between these two potential explanations arises from the fact that they have contrary implications for policy. "Learning-by-doing" acts as an argument in favor of increased concentration among providers, whereas regionalizing procedures in the face of "selective referral" only leads to reduced competition, without any improvement in outcomes.

A handful of studies have attempted to distinguish between these explanations using data at the level of the firm (hospital) by using instruments that are correlated with procedure volume but unrelated to factors that influence patient outcomes. Luft, Hunt and Maerki (1987) use cross-sectional data to

[6] See Luft et al (1990) for a comprehensive review of volume-outcome studies, and Halm et al (2002) for a review of volume-outcome studies related to CABG.

estimate a simultaneous equations model where they use the size of the hospital, teaching affiliation, and the number of appendicitis procedures as instruments in the equation predicting volume. However, to the extent that the quality of different procedures in a hospital may be related or that a teaching hospital may attract sicker patients, it is not immediately clear why these instruments are excluded from the outcome equation.

A few recent studies (Gowrisankaran, Town and Ho (2006), Picone, Trogdon and Trollis (2005)) use predicted volume as an instrument for actual procedure volume (at the hospital level), where predicted volume is estimated from a multinomial hospital choice model based on distance and basic hospital characteristics. The identification in these models comes from the fact that the location of patients relative to hospitals is a key determinant of hospital volume and is assumed to be exogenous to outcomes. These studies conclude in the favor of a strong learning-by-doing effect for hospitals performing cardiac procedures. While working with physician level data (as this paper does), this instrument is not as useful because one needs to sort out the patient's choice of a hospital from her choice of a physician. Further, these studies cannot test for the presence of any aspects of specificity in learning-by-doing.

A few other studies have used panel data to control for unobserved determinants of hospital quality by including hospital fixed effects in their regressions. Examples of such studies include Ho (2002), Farley and Ozminkowski (1992) and Hamilton and Hamilton (1998). Their estimates, obtained using within-hospital variation in volumes, suggest that hospital volume leads to better outcomes. However, the possibility that changes in within-hospital volume are driven by unobservable elements of hospital quality still cannot be ruled out.

Analyses of volume-outcome effects using data at the surgeon level have tended to be largely correlational – i.e., these studies assume the presence of a positive correlation between provider volume and outcomes as indicative of a learning-by-doing effect. Examples of such studies include Hannan et al (1991) and Hughes et al (1987). Failure to account for selective referral in these studies may bias estimated effects of volume on outcome.

A further problem with correlational studies (that use OLS regression models of patient outcome on provider volume) is that one cannot even sign the direction of the bias on the volume coefficient. The presence of selective referral should serve to make the OLS estimate larger in absolute value when compared to the IV estimate which parcels out this effect. However, if patients who are attracted to high quality surgeons are more ill in ways that are unobservable, it could lead to an increase in bad outcomes for the high quality surgeons, and result in OLS coefficients underestimating the true effect of learning-by-doing (in absolute terms). The only way to determine the direction of bias is by comparing OLS estimates with those obtained by IV regression. Section 2.6 presents results from such a comparison.

Huckman and Pisano (2006) were the first to examine the important question of firm-specificity of performance by using data on cardiac surgeons in the state of Pennsylvania, from 1994-1995. In particular, they look at whether a surgeon's experience (measured by procedure volume) at one hospital translates into better outcomes for her patients at other hospitals she operates in. However, they do not control for surgeon fixed effects and fail to account for endogeneity of surgeon volume in their empirical specifications. Their estimates are thus identified off levels implying that unobservable surgeon-specific factors might bias their results. Based on their assumptions, they find a strong learning-by-doing effect and evidence to support firm-specificity of surgeon performance. While the focus of this study is to document the presence of learning effects among surgeons using robust empirical methods, the study setting allows me to make some inferences about the specificity of these effects as well.

Previous studies of individual (surgeon) learning in health care suffer from limits imposed by data as well. Either the panel does not track surgeons over a long enough time period (e.g. Huckman and Pisano (2006)) or the data does not track individual surgeons but only identifies the proportion of procedures performed by high and low volume surgeons (e.g. Hughes et al (1987)). The dataset used in this study overcomes both these limitations and tracks individual surgeons over a six year period.

In summary, the main contributions of this study lie in using detailed micro-level data and robust empirical techniques to shed light on the process by which individuals within organizations attain (or maintain) proficiency at their tasks. In doing so, I aim to contribute to the literatures on organizational learning, human capital management and health economics. In addition, the basic idea underlying the instrument, that of exitor behavior affecting incumbents, is applicable in other settings: for example, one can examine the effects of a firm closure on other firms in a market. The study thus makes a useful contribution to the applied econometrics literature as well.

2.3 Data and Research Setting

2.3.1 Selecting a Candidate Procedure

I study the impact of learning-by-doing on quality for surgeons who perform Coronary Artery Bypass Grafts (CABG). Developed in the late 1960s[7], CABG is a risky and invasive surgical procedure that is normally performed on patients with severe or multiple narrowing of the coronary arteries. It is one of the ways[8] in which Coronary Artery Disease, one of the leading causes of death in the US[9], is treated. The procedure involves bypassing a blocked (or narrowed) segment of a heart artery by using a graft from the arm, leg or chest. Hospitals typically have dedicated operating

[7] Source: Website of the American Heart Association
[8] Other treatments include medication and angioplasty
[9] Source: National Center for Health Statistics, <http://www.cdc.gov/nchs/fastats/lcod.htm>

rooms, fitted with specialized equipment and manned by dedicated technicians, in which the procedure is performed.

The referral process for CABG normally works as follows: a patient experiencing chest pains or shortness of breath starts by visiting a primary care physician who may then refer her to a cardiologist for further treatment and evaluation. The cardiologist evaluates the patient's medical history and symptoms and may perform a cardiac catheterization, a procedure that indicates how well blood is flowing through the vessels that supply the heart muscle. If the catheterization shows abnormal results, the patient is treated according to the extent of blockage in the arteries. Mild to medium blockages are treated using medication or angioplasty[10], while severe cases are referred to a surgeon for bypass surgery.

There are multiple reasons why CABG was selected as a candidate procedure for this study. First, it is a procedure that can be performed only by highly trained and specialized surgeons whose abilities are perceived as being crucial to the quality of care. Given the key role played by the surgeon, it is important to study the factors that affect surgeon performance. Second, CABG is a fairly common procedure with over 470,000 surgeries being performed in the US in 2004.[11] It accounts for ~4-5% of total health expenditure in the US[12], and is hence important in its own right. Third, the procedure has been extensively studied in the health economics literature with the result that there is a commonly accepted and readily available measure of outcomes: in-hospital mortality.[13]

---

[10] This procedure is performed by an interventional cardiologist.
[11] Source: 2004 National Hospital Discharge Survey
[12] Ibid.
[13] An alternate measure of mortality used by some volume-outcome studies is 30-day mortality, defined as death occurring within 30 days of admission. The Florida AHCA data only identifies in-hospital mortality, while Medicare data files (which report 30-day mortality rates) typically do not include physician identifiers. So, I proceed with in-hospital mortality as my measure of patient outcome.

Finally, since CABG is performed by surgical specialists, it is likely that a sizeable proportion of patients are referred to the appropriate hospital or surgeon (by their cardiologist) on the basis of provider quality – implying that "selective referral" could play an important role. Therefore, it becomes all the more important to not rely on correlational studies of volume-outcome while estimating learning-by-doing effects.

2.3.2 Data

The primary dataset for this study comes from the Hospital Inpatient Data Files provided by the Florida Agency for Health Care Administration (AHCA) for the years 1998-2003. This data is comparable to patient level discharge data provided by the Health Care Utilization Project (HCUP) and the California Office of Statewide Health Planning and Development (OSHPD). Specifically, AHCA provides information about each discharge including the hospital, physical and demographic characteristics of the patient (e.g. age, sex, payer information), and a comprehensive list of primary and secondary diagnoses. In addition, the data identify the license number of the operating surgeon, enabling me to track procedure volume of individual surgeons over a six year time frame.

I also make use of two secondary datasets that provide me with information on hospital and physician characteristics. I obtain hospital level data from the American Hospital Association (AHA) Annual Survey of Hospitals for the years 1998-2003.[14] I obtain detailed information about each surgeon's training (e.g. medical school/residency program trained at, year of graduation) and draw inferences about the physician's age by linking the license number to an online database provided by the Florida Department of Health.[15] To protect confidentiality, I do not present any physician-specific information.

---

[14] Since the AHCA dataset does not provide AHA ids for hospitals, I link it to the AHA data using the Medicare (HCFA) id for each hospital and was able to obtain perfect matches for all but one hospital.
[15] This information is available at <http://ww2.doh.state.fl.us/irm00praes/praslist.asp>

I restrict attention to surgeons who appear more than once in the panel, and perform at least 5 procedures over the entire panel in order to exclude the effects arising from unrepresentative surgeons, e.g. an Emergency Room surgeon who may perform the occasional surgery. I also exclude hospitals that appear in the bottom one-half percentile of the distribution of hospital procedure volume. The empirical results are robust to these sample restrictions.

2.4 Using Surgeon Exit to Identify Exogenous Shocks to Procedure Volume

This paper uses an instrumental variables approach to determine the extent to which increased procedure volume for surgeons translates into better outcomes for their patients. An ideal instrumental variable in this case should have the following properties: it should help explain variation in surgeon procedure volume (the endogenous predictor), and have no causal relationship with surgeon quality except through its effect on procedure volume. One can argue that concern about "selective referral" is mitigated at the surgeon level – this would hold true if patients only choose hospitals (and not individual surgeons) based on quality. However, to the extent that patients are typically referred to providers by other physicians[16] for procedures like the one studied in this paper (CABG), the possibility that higher quality surgeons attract more patients cannot be ruled out.

This paper proposes the use of surgeon exit as an exogenous identifier. I use the term "exit" to refer to any instance where a surgeon stops performing CABG surgeries in Florida.[17] A surgeon may exit because of old age (retirement), death, termination of employment, or relocation to a different state.[18] The rationale is as follows: once a surgeon exits a hospital for exogenous reasons, her procedure volume gets redistributed among non-exiting surgeons at the hospital, giving their

[16] A patient first visits a cardiologist who may then refer her to a cardiac surgeon.
[17] I consider a surgeon to have exited when she stops performing procedures across all hospitals in Florida. I use this definition in order to rule out cases where a surgeon may exit a particular hospital but enter a nearby hospital, thereby taking patients with her.
[18] The data does not identify the cause of exit. However, I use data on the surgeon's age to make some inferences about possible cause of exit, as described in Section 4.1.

procedure volumes a positive shock. If this shock to surgeon volume is not correlated with unobservable determinants of changes in surgeon quality, the use of exit as an identifier is valid. The instrumental variable used in the regressions is defined as a function of the volume of the exiting surgeon(s). I first provide a formal definition of the instrument before discussing some of the underlying identifying assumptions in detail.

2.4.1 Defining the Instrument

I define surgeon exit as follows: a surgeon exits a hospital if her CABG procedure volume at that hospital drops to zero, and she performs at least 5% of the hospital's CABG procedures in the year preceding exit. The latter part of the definition ensures that surgeons who perform the occasional CABG procedure are not counted amongst exiting surgeons. I consider exit to have taken place only when a surgeon stops performing procedures across all hospitals she was operating in.[19] I further restrict the set of exitors to surgeons aged 55 and over.[20] Doing so increases the probability that a surgeon exited the data because she was retiring and mitigates concerns about endogeneity of exit. In section 2.6.4, I test the robustness of the empirical results by slightly varying the definition of exit and find little change in the conclusions.

The volume of the exiting surgeon is simply calculated as the number of procedures performed by the surgeon in year *t-1*, where *t* denotes year of exit. I assume that the impact of surgeon exit is felt on staying surgeon volumes for one year, i.e. if a surgeon exits in year *t*, then existing surgeons in a hospital experience an exogenous shock to their procedure volumes in year *t* only. The timing of the model works as follows: if a surgeon exits (i.e. stops performing procedures) in year *t*, the non-

[19] In a majority of cases, a surgeon exits all hospitals simultaneously and that year is taken to be the year of exit. In a few instances, a surgeon who works across multiple hospitals may exit one hospital but continue working in another. In such cases, I consider the surgeon to have exited only when she stops performing procedures altogether, and the latter year is taken as the year of exit from the data.

[20] The data on surgeon characteristics does not contain the age of the surgeon, but does contain the year in which the surgeon graduated from medical school. I infer the surgeon's age using this information.

exiting surgeons at that hospital experience a positive shock to volume in year *t*. This volume shock translates into better patient outcomes for these surgeons in year *t+1*.

While one can use the data to identify instances of surgeon exit, there is no way of using the data to determine how exit volume[21] is allocated among staying surgeons. There are many ways in which patients who would have visited the exiting surgeon may be redistributed across non-exiting surgeons. I discuss three different methods of allocating exit volume across surgeons and test my main specifications under all three scenarios.

Under the first allocation rule, patients of the exiting surgeon are assumed to be allocated equally among all the remaining surgeons at the hospital. The instrument is then computed as the sum of the procedure volumes of all exiting surgeons in that hospital. In other words, all surgeons working at a hospital face the same shock to volume (equal to the above sum of exit volumes), irrespective of their shares prior exit.

One demerit of the instrument as defined above is that it exhibits no variation across surgeons within a hospital in a particular year. In other words, a high volume surgeon stands to gain as much from exit as a low volume surgeon. The second allocation rule addresses this issue by allocating exit volume to staying surgeons in proportion to their share of procedure volume within the hospital in year *t-1*, where *t* denotes the year of exit. To the extent that exit volume is allocated to staying surgeons in proportion to their current shares, this volume shock is exogenous to changes in surgeon quality.[22] In specifications that do not have surgeon fixed effects, however, the identification assumption may be violated.

---

[21] I use the term exit volume to refer to the number of procedures performed by an (exiting) surgeon at a hospital the year before exit.

[22] The main specifications include surgeon fixed effects, implying that we look at changes in surgeon quality, and not the absolute level.

The third allocation rule also leads to variation in the instrument across surgeons within a hospital and also guards against endogeneity. Under this rule, I assign exit volume to each staying surgeon on the basis of the extent of patient zip code overlap with the exiting surgeon. To illustrate the methodology, consider a hospital with three practicing surgeons in year *t-1*, one of whom (surgeon A) exits in year *t*. Table A1 presents the breakdown of the exiting surgeon's volume by the zip code of the patients treated. Rows 2 and 3 of Table A1 present the distribution of patients across those zip codes operated upon by the other (non-exiting) surgeons in the hospital.[23]

Since surgeon A exits the data in year *t,* her procedure volume in year *t-1* is assigned to the other surgeons in the hospital in year *t*. Table A2 presents the allocation of A's exit volume to the non-exiting surgeons in the year post exit. Note that in Column 1, surgeon C is allocated no patients from zip code 60201 as she does not treat any patients from that location. As a result, the zip code-level exit volume of 10 patients is entirely allocated to surgeon B. In Column 2, the zip code-level exit volume (of 8 patients) is equally divided amongst both staying surgeons as they both treat patients from that zip code in *t-1*. Note that division of exit volume in this zip code is independent of surgeon market share (in that zip code) – this allocation rule tries to ensure that the additional experienced gained by a surgeon from exit is independent of her current procedure volume, which may reflect unobserved components of quality. Finally, patients in non-overlapping zip codes (zip code 60110 in the example above) are allocated equally across all non-exiting surgeons.

[23] Note that the non-exiting surgeons may draw patients from other zip codes as well – this table lists patient zip codes of the exiting surgeon only.

Table A1. Zip code breakdown of surgeon procedure volume, year *t-1*

	Number of patients treated by surgeon in zip code, *year t-1*		
	60201	**60031**	**60110**
Surgeon A	10	8	6
Surgeon B	3	5	0
Surgeon C	0	1	0

Table A2. Calculating exogenous volume shock, year *t*

	Allocation of exit volume, *year t*			
	60201	**60031**	**60110**	**Total**
Surgeon B	10	4	3	17
Surgeon C	0	4	3	7

As mentioned earlier, there is no way of empirically determining the exact mechanism through which patients of the exiting surgeon are allocated. In the remainder of this paper, I present results only using the third allocation rule: the zip code overlap method. The results[24] using the two other

[24] Available upon request

allocation rules discussed earlier were qualitatively similar to the main results described in Section 2.6.

2.4.2 Why is Surgeon Exit a Plausible Instrument?

I now discuss some of the underlying assumptions in using surgeon exit to identify exogenous changes in surgeon procedure volumes. I test the validity of these assumptions using the data, and include these results in a later section.

The correlation between the procedure volumes of staying surgeons and the volume of exiting surgeons (i.e. the relevance of the instrument) could be mitigated under some conditions. The first possibility is that the hospital hires new surgeons who take over the caseload of exiting surgeons. In that case, the procedure volume of non-exiting surgeons is unaffected by exit. However, the data show that surgeon exit is accompanied by entry (the same year) in only 12% of the cases. Further, an entering surgeon performs only 21 procedures at a hospital in the year of entry, compared to an exiting surgeon who performs 45 procedures on average. Taken together, these facts imply that entry of new surgeons should not affect instrument relevance significantly.

The second possibility is that the procedure volume of the exiting surgeon is "lost" to the hospital, i.e. patients who would have visited the exiting surgeon now choose a different hospital instead of visiting another surgeon in the same hospital. However, the data clearly indicate that there is very little variation in hospital volume over time; hospital volume decreases by 2% the year of exit, compared to a 2% annual increase over all years. This implies that surgeon exit does not lead to a significant decrease in a hospital's annual procedure volume.

I now turn to the question of instrument exogeneity. The main specifications include surgeon fixed effects implying that the dependent variable is the change in quality of the surgeon. The identifying assumption I make here is that the cause of surgeon exit is not directly related to changes in (non-exiting) surgeon quality. This assumption ensures that the instrument (which is a function of exiting surgeon volume) is uncorrelated with unobservable determinants of changes in staying surgeon quality. An instance in which this assumption would be violated is if a surgeon exited because she believed her rivals' quality was going to improve in the future. Consider a scenario in which the exiting surgeon's volume is determined by her quality relative to the quality of the staying surgeon. In other words, an exiting surgeon might have higher procedure volume if her rival surgeons were of poor rather than high quality compared to her own quality. This would induce correlation between exit volume and changes in (staying) surgeon quality, thereby confounding identification. I test this assumption using the data and provide the details of this test in Section 2.6.

2.5 A Robust Empirical Model of Learning-by-Doing

2.5.1 Instrumenting for Total Surgeon Experience

In order to estimate the extent of surgeon learning-by-doing, I model surgeon quality as a function of various characteristics of the patient being treated, of the hospital she is treated at, and of the surgeon performing the procedure, including surgeon experience. Specifically, I estimate an instrumental variables regression (using the Stata module *xtivreg*) where observations are at the level of the patient and the endogenous predictor, surgeon experience, is instrumented for. Equations (1.0) and (1.1) represent the first[25] and second stages, respectively, in the instrumental variables estimation procedure. The variables and notation are explained in detail below, starting with the second stage regression equation. In all equations, *i* indexes the patient, *h* indexes the hospital, *p* indexes the surgeon and *t* indexes the time period of observation.

$$\begin{aligned}(Physvol)_{p,t-1} &= \alpha_0 + \alpha_1 * (Exitvol)_{p,t-2} + \alpha_2 * X_{i,p,h,t} + \alpha_3 * \theta_p + \alpha_4 * \mu_h \\ &\quad + \alpha_5 * \lambda_{h,t} + \alpha_6 * (Year)_t + \upsilon_{p,t-1}\end{aligned} \tag{1.0}$$

[25] The unit of observation for the first stage regression is the patient (as in the second stage). However, I have suppressed the patient index *i* in the notation for ease of exposition.

$$(Outcome)_{i,p,h,t} = \beta_0 + \beta_1 * (Ph\hat{y}svol)_{p,t-1} + \beta_2 * X_{i,p,h,t} + \beta_3 * \theta_p + \beta_4 * \mu_h + \beta_5 * \lambda_{h,t} + \beta_6 * (Year)_t + \varepsilon_{i,p,h,t} \tag{1.1}$$

Second Stage Regression

Dependent variable: The dependent variable in the second stage regression is a measure of surgeon quality. Since analyses are conducted at the level of the patient, I use in-hospital patient mortality as the primary measure for surgeon quality. The main advantage of using in-hospital mortality as a measure of outcome is that there is very little chance of miscoding. In addition, there is sufficient variation in outcomes across surgeons and hospitals which allows me to estimate learning effects with precision. However, it may not completely reflect the health status of the patient post surgery. The dependent variable $\text{Outcome}_{i,p,h,t}$ is a binary variable that is set to one if patient *i* died as a result of a CABG procedure performed by surgeon *p* in hospital *h* at time *t*.

Independent variables – surgeon experience: The primary predictor of interest measures the experience of the surgeon performing the procedure. I proxy for surgeon experience by the total number of CABG procedures performed by the surgeon the previous year, $Physvol_{p,t-1}$. This measure is aggregated across all hospitals the surgeon operates at. In eq. (1.1), $Phy\hat{s}vol_{p,t-1}$ denotes the predicted value from the first stage regression, eq. (1.0). The coefficient β_1 tells us the effect of surgeon experience on outcomes, purged of any possible biases arising from endogeneity. A negative sign on β_1 will act as evidence in favor of the learning-by-doing hypothesis. I let surgeon volume enter the specification in a linear fashion above.[26]

[26] In an alternate specification, I used the square root of procedure volume as the measure for volume in order to incorporate nonlinearity. The linear specification was found to have a better fit (in terms of R^2) so I proceed with that for the rest of the specifications. Models using the square root of volume yielded similar conclusions.

While prior theoretical and empirical research on learning-by-doing model experience as having a cumulative effect, I use recent volume to measure surgeon experience mainly because I observe cumulative volume only for a subset of surgeons in the data.[27] The concern over using recent procedure volume instead of cumulative procedure volume is mitigated to an extent by recent research studies (Gowrisankaran, Town and Ho (2006), Gaynor, Seider and Vogt (2005)) that find a significant amount of organizational forgetting among hospitals performing CABG procedures. This implies that experience from the immediate past has a greater impact on outcomes when compared to experience from further before. As a robustness check, I estimate the model using the number of procedures performed by the surgeon in the last *two* years as a proxy for surgeon experience. I also repeat the analysis on a sample that contains only those surgeons who finish their residency training in 1998 or later. In this specification, I use cumulative volume as the proxy for surgeon experience.

I include an interaction term in order to allow for a nonlinear effect of procedure volume on mortality. Specifically, I interact the lagged surgeon volume term, $Physvol_{p,t-1}$, with an indicator for whether the surgeon was a high volume surgeon the previous year.[28] I define a high-volume surgeon as one performing more than 200 procedures that year. The coefficient on the interaction term can be used to infer whether high volume surgeons benefited more (or less) from an exogenous change in experience, when compared to low volume surgeons.

Independent variables – surgeon characteristics: In order to isolate the effect of surgeon experience on outcomes, I control for other characteristics of the surgeon that determine quality. In the main specifications, I do this by including a vector of surgeon fixed effects (θ_p) in the regression that account for the effect of time-invariant surgeon characteristics (e.g. sex, training etc. and

[27] Specifically, the data allow me to track all procedures only for surgeons who finish their residency training in 1998 or later.
[28] One concern here, is of course, the potential endogeneity of the interaction term. The use of an indicator for a high volume surgeon (as opposed to using the actual experience of the surgeon) should lessen this concern to an extent.

unobservables that are fixed with time) on patient outcomes. I also run a specification that excludes surgeon fixed effects[29]; in this model, I control for the following surgeon characteristics: age, sex and whether the surgeon trained at a foreign medical school.

Independent variables – hospital characteristics: I include a vector of hospital fixed effects (μ_h) to control for systematic intrinsic quality differences across hospitals.[30] Following Ho (2002), I also include a vector of hospital characteristics ($\lambda_{h,t}$) that reflect the hospital's scale, staffing levels and service makeup. These include: the number of general and cardiac care beds, number of FTE registered nurses and licensed practical nurses, number of high-tech services offered (MRI, CT scan, PET scan etc) and indicators for medical school affiliation and having an approved residency program.[31]

Independent variables – patient characteristics: The vector $X_{i,p,h,t}$ includes patient characteristics (e.g. the number of co-morbidities, patient age, sex, an indicator for whether the patient has had a prior CABG), all of which are expected to have an impact on patient mortality. The patient characteristics I include in the regression are: patient age categories, sex, the Charlson co-morbidity index[32], concurrent angioplasty, cardiogenic shock, prior CABG surgery, congestive heart failure, hypertension and an indicator for a heart attack. All specifications also contain year indicators to control for the effect on outcomes of changes in technology that are not captured by other predictors.

---

[29] I do this mainly to facilitate comparison with earlier surgeon level volume outcome studies, that do not include surgeon fixed effects in their specifications.

[30] Note that hospital fixed effects are not perfectly collinear with surgeon fixed effects as surgeons may work across multiple hospitals in the same year.

[31] Since I also include hospital fixed effects in the regression, the explanatory power of these variables is rather limited. An F-test found these variables to have some joint predictive power so I include them in the regression.

[32] The Charlson co-morbidity index reflects the cumulative increased likelihood of one year mortality arising from different categories of co-morbidity such as cancer, diabetes, AIDS etc. The higher the score, the worse the condition of the patient.

First Stage Regression

In the first stage regression (eq. (1.0)), surgeon volume is estimated as a function of the instrument and all exogenous predictors from the second stage. Note that a staying surgeon's procedure volume in year *t-1* is affected by surgeons who exit in year *t-1*. Thus, the instrument allocates exiting surgeons' procedure volume in year *t-2* to staying surgeons in year *t-1*. I use the sum of the exit volumes allocated to the staying surgeon across all hospitals to instrument for past procedure volume of the staying surgeon. The effect of exit on the procedure volume of staying surgeons is assumed to last for a year. Based on the discussion in Section 2.4, the coefficient α_1 is expected to have a positive sign in eq. (1.0).

2.5.2 Instrumenting for Specific Aspects of Surgeon Experience

The regression equations for exploring specificity of surgeon experience are set up in a similar manner. I estimate a patient-level instrumental variables regression where the dependent variable is, as before, a measure of surgeon quality (patient outcome). The key difference is that I now use a hospital-specific measure of surgeon experience as the primary predictor. The variable $Physhospvol_{p,h,t-1}$ measures the procedure volume of surgeon *p* at hospital *h* in the year *t-1*, while $Physhospvol_{p,-h,t-1}$ measures the procedure volume of the same surgeon across all other hospitals (apart from hospital *h*) in *t-1*. In Eq. (1.4), the variables $Physh\hat{o}spvol_{p,h,t-1}$ and $Physh\hat{o}spvol_{p,-h,t-1}$ refer to the predicted values of these variables from the first stage regressions.

$$\begin{aligned}(Physhospvol)_{p,h,t-1} &= \alpha_0 + \alpha_1 * (Exitvol)_{p,h,t-2} + \alpha_2 * (Exitvol)_{p,-h,t-2} + \alpha_3 * X_{i,p,h,t} \\ &+ \alpha_4 * \theta_p + \alpha_5 * \mu_h + \alpha_6 * \lambda_{h,t} + \alpha_7 * (Year)_t + \upsilon_{p,h,t-1}\end{aligned} \tag{1.2}$$

$$\begin{aligned}(Physhospvol)_{p,-h,t-1} &= \gamma_0 + \gamma_1 * (Exitvol)_{p,h,t-2} + \gamma_2 * (Exitvol)_{p,-h,t-2} + \gamma_3 * X_{i,p,h,t} \\ &+ \gamma_4 * \theta_p + \gamma_5 * \mu_h + \gamma_6 * \lambda_{h,t} + \gamma_7 * (Year)_t + \eta_{p,-h,t-1}\end{aligned} \tag{1.3}$$

$$(Outcome)_{i,p,h,t} = \beta_0 + \beta_1 * (Phys\hat{h}ospvol)_{p,h,t-1} + \beta_2 * (Phys\hat{h}ospvol)_{p,-h,t-1} + \beta_3 * X_{i,p,h,t} + \beta_4 * \theta_p + \beta_5 * \mu_h + \beta_6 * \lambda_{h,t} + \beta_7 * (Year)_t + \varepsilon_{i,p,h,t} \quad (1.4)$$

Correspondingly, I use hospital-specific instruments for surgeon procedure volume in the first stage regressions (1.2) and (1.3). $Exitvol_{p,h,t-2}$ is calculated as the exit volume allocated (using the extent of zip code overlap) to surgeon *p* at hospital *h* in *t-2,* while $Exitvol_{p,-h,t-2}$ is calculated as the sum of exit volumes allocated to surgeon *p* across all other hospitals (other than *h*) in *t-2*.

If human capital acquired by learning-by-doing were firm-specific, one would expect additional surgeon experience to have a greater benefit on patient outcomes at the hospital where the experience was gained. In other words, if β_1 and β_2 are both negative, and furthermore, β_2 is smaller in absolute value than β_1, the implication is that procedure volume at other hospitals does impact surgeon quality at the hospital under consideration, but not as much as procedure volume at the same hospital. This would be evidence in support of firm-specificity. A stronger form of firm specificity, in which experience is completely non-transferable across firms, would have β_2 equal to zero (or statistically indistinguishable from zero). On the other hand, if performing additional procedures (irrespective of hospital) adds to a surgeon's general human capital, one would expect β_1 and β_2 to be statistically indistinguishable from each other.

I test for task-specificity of surgeon human capital by using a specification similar to the one described above. Instead of using hospital-specific measures of surgeon volume, I now use task-specific measures. Specifically, I divide the total number of cardiac procedures performed by each surgeon into the number of CABG and the number of non-CABG procedures. I construct analogous versions of the instrument and compare the coefficients on the volume terms to make inferences on task-specificity of human capital.

2.6 Results

Before discussing the regression results, I present some patterns in the raw data. The study uses data on a total of 385 CABG surgeons working at 65 hospitals in Florida over 6 years, with a total volume of 160,210 procedures. The number of surgeons in each year increases from 224 in 1998 to 265 in 2003, while the number of hospitals increases from 57 to 65 over the same time period. Table 2.1 presents descriptive statistics of some of the key variables used in the specification. These statistics are calculated from patient-level data. The dependent variable, patient mortality, has a mean of 3.82% in the data, which is in line with mortality rates observed in other studies. The patient population is composed of a majority of males (~70%) who were over 65 years old on average. Around 6% of the population had had a previous CABG surgery.

The table also presents surgeon and hospital characteristics (averaged over patient level data). CABG surgeons in the data were overwhelmingly male (almost 98%) and a majority of them graduated from a US medical school (~83%). The mean age of surgeons in the data is almost 40. The average surgeon performs almost 175 procedures annually, across all hospitals. This number drops to 143 when hospital-specific volume is considered. These numbers do not reflect the true distribution of surgeon procedure volume, where one can see a bunch of low-volume surgeons. The average hospital employs around 6.5 surgeons. More than 40% of surgeons work across multiple hospitals during the same time period. The mean for the instrument is rather low (2.95) because of the number of cases in which no surgeon exit is recorded in a hospital (the instrument is assigned a value of zero in this case). In all, I record 33 instances where a surgeon exits the data. 230 non-exiting surgeons in the data are affected at some point in time by surgeon exit and experience a positive shock to their volumes.

Table 2.1: Summary Statistics for Key Variables

	Mean	Std. Deviation	Number of Obs.
Dependent Variable			
Mortality Rate	3.82%	19.17%	160210
Volume Measures			
Total Surgeon Volume, prior year	174.6	82.9	129554
Hospital-Specific Surgeon Volume, prior year	142.9	86.3	126291
Hospital volume	677.4	473.3	160210
Patient Characteristics			
Age	67.5	10.6	160210
% Female	29.4	45.6	160210
Charlson Index	.88	.97	160210
Prior CABG surgery	.06	.24	160210
Concurrent PTCA	.03	.16	160210
Heart Failure	.21	.41	160210
Cardiogenic shock	.02	.14	160210
Hypertension	.01	.10	160210
Hospital Characteristics			
Number of cardiac care beds	23.9	30.6	130371
Number of high-tech services	2.4	.82	130371
Number of FTE Registered Nurses	160.7	.16	160210
Number of FTE Licensed Practical Nurses	86.3	.41	160210
Surgeon Characteristics			
Years since Medical School	21.5	7.8	160210
% Foreign Medical School	17.4	.38	160210
Instrument			
Exit Volume	2.95	10.3	129554

Note: All statistics are calculated using the patient as the unit of observation. The sample contains all patients treated with CABG surgery in Florida for the years 1998-2003. The sample excludes a few low-volume surgeons and hospitals (refer section 2.3 for details). The discrepancy in the number of observation for surgeon and hospital characteristics is due to missing data. In the case of the volume measures, the discrepancy is because these measures are lagged by one year.

2.6.1 Testing the Identification Assumption

The main identification assumption behind using exitor volume as an instrument is that it is uncorrelated with unobserved determinants of changes in staying surgeon quality. I validate this assumption by testing if the amount of exit volume faced by a surgeon is determined by her quality in the previous period. This reduced form regression is performed at the level of the surgeon-year and uses the instrument as the dependent variable and the lagged risk-adjusted surgeon-level mortality rate[33], $(Mortrate)_{p,t-1}$, as the primary predictor. I also include surgeon characteristics ($\psi_{p,t}$) – age, sex, whether the surgeon is foreign trained – and surgeon-level means of all the patient characteristics described earlier ($X_{p,t}$) as predictors.

$$(Exitvol)_{p,t} = \beta_0 + \beta_1 * (Mortrate)_{p,t-1} + \beta_2 * X_{p,t} + \beta_3 * \psi_{p,t} + \beta_4 * \lambda_{h,t} + \beta_5 * (Year)_t + \xi_{p,t} \tag{1.5}$$

The results, reported in Table 2.2, indicate that the coefficient on the primary predictor, lagged surgeon quality, is statistically indistinguishable from zero (*p=.984*). This suggests that the amount of exit volume faced by a surgeon is independent of her prior quality, and validates the main identification assumption in the paper.

2.6.2 Do Surgeons Learn From Experience?

Table 2.3 provides the main results from first-stage regressions of lagged surgeon procedure volume (across all hospitals) on the instrument, along with all other exogenous predictors from Stage 2. Columns 1 and 2 present the results without and with surgeon fixed effects. The coefficient on the instrument is positive and highly significant indicating that exit volume faced by the surgeon is

[33] In order to compute risk adjusted mortality rate for each surgeon, I run a patient-level logit model where the dependent variable equals one if the patient died, and the predictors include all patient characteristics and a vector of year fixed effects. I then sum up the predicted probability of mortality for all patients across a surgeon and divide by the number of patients treated by that surgeon.

strongly correlated with actual procedure volume, with t-statistics of 16.62 and 20.43 respectively. The magnitude of the coefficient falls (.61 vs .91) on including surgeon fixed effects, implying that between-surgeon variation was driving some of the effects in column 1. The F-statistic[34] for the instrument significance is substantially larger than the typical recommended thresholds[35], validating use of the instrument.

Columns 1 through 6 of Table 2.4 present the results from specifications that estimate the extent of surgeon learning-by-doing. Columns 1, 3 and 5 are estimated without surgeon fixed effects but include a vector of surgeon characteristics.[36] Columns 2, 4 and 6 include surgeon fixed effects to capture the effect of surgeon-specific time invariant factors on mortality.

I present results from OLS regression models (that treat surgeon volume as exogenous) of patient mortality on surgeon volumes in columns 1 and 2, in order to facilitate comparison with existing volume-outcome studies, and to demonstrate the need for instrumenting for surgeon volume. While the first column shows evidence for an inverse relationship between surgeon volume and patient mortality, this cannot be treated as evidence for learning-by-doing because this coefficient contains the influence of "selective referral" as well. Moreover, the coefficient drops drastically in magnitude (and becomes statistically insignificant) once surgeon fixed effects are included, implying that the negative coefficient was not a result of learning-by-doing but was being identified off differences across surgeons.

[34] In the case of a single instrument, the F-statistic is simply the squared value of the t-statistic.
[35] As a rule of thumb, Staiger and Stock (1997) recommend a first-stage F statistic of at least ten for an instrument not to be considered weak.
[36] Models that do not contain surgeon fixed effects have standard errors clustered by surgeon.

Table 2.2: Testing the Identification Assumption: Is the Instrument related to Surgeon Quality?

Dependent Variable: Exit volume faced by surgeon

Lagged Surgeon Quality	0.002
	(0.116)
Charlson Index	-1.471**
	(0.588)
Cardiogenic Shock	-2.196
	(4.511)
Concurrent PTCA	1.477
	(3.229)
Hypertension	0.891
	(4.787)
Heart Failure	0.338
	(1.417)
Prior CABG	1.597
	(2.788)
Female	-1.328
	(1.505)
Heart Attack	0.884
	(1.425)
Age	-0.098**
	(0.04)
Hospital Characteristics	Y
Year Fixed Effects	Y
Number of Observations	936

Note: Regression carried out at the surgeon-year level. Standard errors are reported in parentheses.

**** signifies p<.01, ** signifies p<.05 and * signifies p<.1*

Table 2.3: Relationship between Surgeon Volume and Exit Volume (First Stage)

Dependent variable: Surgeon procedure volume, prior year

	(1)	(2)
Exit Volume, prior year	0.914***	0.611***
	(0.055)	(0.03)
Charlson Index	-0.02	0.047
	(0.196)	(0.098)
Cardiogenic Shock	-2.505*	0.516
	(1.36)	(0.683)
Concurrent PTCA	-2.584**	-0.251
	(1.208)	(0.606)
Hypertension	-0.754	1.727*
	(2.018)	(1.012)
Heart Failure	-1.540***	-0.08
	(0.476)	(0.239)
Prior CABG	3.999***	0.672
	(0.823)	(0.413)
Female	-0.645	-0.032
	(0.418)	(0.21)
Heart Attack	-3.776***	0.268
	(0.484)	(0.243)
Age Categories	Y	Y
Hospital Characteristics	Y	Y
Surgeon Characteristics	Y	N
Year Fixed Effects	Y	Y
Hospital Fixed Effects	Y	Y
Surgeon Fixed Effects	N	Y
Number of Observations	107588	107977

Note: Columns 1 and 2 present models without and with surgeon fixed effects, respectively. The difference in number of observations is due to missing data on some surgeon characteristics. Standard errors are reported in parentheses. Models without surgeon fixed effects have standard errors clustered by surgeon.

In order to exclude any influence of the "selective referral" effect, I instrument for surgeon volume and present these results in columns 3 through 6. The result in Column 3 lends strong support to the learning-by-doing hypothesis: the coefficient on volume is negative and significant (*p=.005*). The magnitude of the coefficient implies that performing one additional procedure in the previous year leads to a decline of .05 percentage points in mortality. This represents a decrease of around 1.2% relative to the average value of mortality in the data (.0382), which is indicative of a strong learning-by-doing effect. Column 4 adds surgeon fixed effects to the model, implying that learning effects are now computed only using within-surgeon variation. The coefficient on lagged surgeon volume is still strongly significant (*p=.015*), and the magnitude of the volume coefficient is now slightly larger in magnitude (-.0007 as compared to -.00051 in column 3). The results of a Hausman test (not included here) clearly reject equality between OLS and IV estimates. The coefficient on volume obtained from OLS regression is smaller in magnitude when compared to that obtained from the IV regressions. As discussed earlier, this implies that the type of patients attracted to high quality surgeons tend to be sicker in ways that are unobservable.

In columns 5 and 6, I include an interaction term in order to allow for a nonlinear effect of procedure volume on mortality. The coefficient on procedure volume is still negative and does not change much in magnitude when compared to the earlier specifications (-.00058 vs. -.00051 for the corresponding specification in Column 3). The coefficient on the interaction term, however, is positive and significant in Column 5 and falls just short of statistical significance at the 10% level in the fixed effects specification in Column 6 (*p=.011*). These results indicate that high volume surgeons (defined as those performing 200 procedures a year or above) benefit less from additional experience when compared to their counterparts who do fewer procedures a year.

2.6.3 Is Experience Specific?

Column 1 of Table 2.5 presents results from the model in Section 2.5.3 that tests for the firm-specificity of human capital acquired through learning-by-doing. The estimates reject the hypothesis of firm-specificity. The coefficient on own-hospital volume is negative and significant (*p=.015)* and only slightly larger in magnitude than the coefficient on other-hospital volume, which is also statistically significant (*p=.05*). Further, a t-test is not able to reject equality of the coefficients (*p=.729*) on the two variables measuring procedure volume (at own and other hospitals), indicating that the benefits of procedure experience are portable across hospital settings. This finding is in contrast to the findings of Huckman and Pisano (2006) who find that performance of cardiac surgeons is firm-specific.

To understand the implication of this result, we need to first examine the exact mechanism through which experience affects the human capital of surgeon in this setting. As discussed earlier, increased procedure experience helps surgeons increase their familiarity with any complications that may arise. It is reasonable to believe that these are benefits which will boost surgeon quality no matter where the procedure is performed. Firm-specific settings play an important role as well for reasons stated earlier. However, the empirical results seem to suggest that the general aspect of surgeon human capital that is gained from experience outweighs the specific aspect.

Table 2.4: The Effect of Total Surgeon Experience on Patient Outcomes

Dependent Variable: Did the Patient Die?

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) IV	(6) IV
Surgeon Procedure Volume, prior year	-3.9E-05***	-2.31E-06	-5.08E-04***	-7.0E-04**	-5.9E-04***	-7.8E-04**
	(8.96E-06)	(1.78E-05)	(1.79E-04)	(2.89E-04)	(2.08E-04)	(3.13E-04)
Surgeon Volume (prior year)*High Vol					3.64E-04***	2.3E-04
					(1.38E-03)	(1.42E-04)
Charlson Index	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cardiogenic Shock	0.341***	0.343***	0.342***	0.341***	0.342***	0.341***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Concurrent PTCA	0.010***	0.011***	0.009***	0.010***	0.010***	0.009***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Hypertension	0.028***	0.028***	0.027***	0.029***	0.027***	0.029***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Heart Failure	0.030***	0.030***	0.029***	0.030***	0.029***	0.030***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Prior CABG	0.029***	0.029***	0.031***	0.029***	0.031***	0.029***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Female	0.014***	0.015***	0.014***	0.014***	0.014***	0.014***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Heart Attack	0.005***	0.005***	0.004**	0.006***	0.004***	0.005***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Age Categories	Y	Y	Y	Y	Y	Y
Hospital Characteristics	Y	Y	Y	Y	Y	Y
Surgeon Characteristics	Y	N	Y	N	Y	N
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Hospital Fixed Effects	Y	Y	Y	Y	Y	Y
Surgeon Fixed Effects	N	Y	N	Y	N	Y
Number of Observations	107588	107977	107588	107977	107588	107977

Note: Columns 1 and 2 present OLS models without and with surgeon fixed effects, respectively. Columns 3 through 6 present IV estimates. Standard errors are reported in parentheses. Models without surgeon fixed effects have standard errors clustered by surgeon. The slight discrepancy in the number of observations across models is due to missing data on surgeon characteristics.
**** signifies p<.01, ** signifies p<.05 and * signifies p<.1*

An interesting implication of this result is on the division of rents between surgeons and hospitals in this setting. The finding that surgeon performance is portable across hospitals suggests that surgeons should end up capturing any rents that may accrue from this relationship. Surgeons are not hospital employees, but typically operate as freelance agents who enter into contractual relationships with hospitals. Further, the law forbids hospitals from paying physicians with the objective of directing the physician's patients towards the hospital. One way in which surgeons may capture some of these rents is by forming their own specialty hospitals along with the help of outside investors. Since learning is portable, high quality surgeons are able to leave a hospital but maintain their skills while starting a new practice.

In Column 2, I present results from the specification testing for task-specificity of human capital. The coefficients on both volume terms (CABG procedure volume and non-CABG procedure volume) are negative and statistically significant at the 5% level. However, the CABG volume coefficient is much larger in magnitude (-.00046 vs. -.00009).[37] These estimates indicate that there is some benefit of performing non-CABG procedures on CABG outcomes, but the impact of this experience is smaller than the impact of experience performing CABG procedures. I interpret this as evidence in favor of some degree of task-specificity of human capital. The main implication of this result is that there seems to be some economies of scope across procedures for surgeons. This provides one explanation as to why surgeons do not specialize in performing just one particular procedure.

[37] A t-test rejects equality of the two coefficients at the 5% level.

Table 2.5: Testing for Specificity of Experience

	Dependent Variable: Did the patient die?	
	(1)	**(2)**
Surgeon Procedure Volume, prior year, own hospital	-5.73E-04**	
	(2.36E-04)	
Surgeon Procedure Volume, prior year, other hospitals	-5.12E-04**	
	(2.65E-04)	
Surgeon Procedure Volume, prior year, CABG		-4.64E-04**
		(2.28E-04)
Surgeon Procedure Volume, prior year, non-CABG		-9.17E-05***
		(2.53E-05)
Patient Characteristics	Y	Y
Hospital Characteristics	Y	Y
Surgeon Characteristics	N	N
Year Fixed Effects	Y	Y
Hospital Fixed Effects	Y	Y
Surgeon Fixed Effects	Y	Y
Number of Observations	105066	107977

Note: Column 1 presents results from a test of firm-specificity of learning-by-doing. Column 2 presents results from a test of task-specificity. Standard errors are reported in parentheses.

**** signifies p<.01, ** signifies p<.05 and * signifies p<.1*

2.6.4 Robustness Checks

I test the robustness of the estimates through various specification checks and sample restrictions. Table 2.6 contains results from these specifications. In all specifications, I focus on the coefficient of lagged surgeon volume which is the main coefficient of interest. In column 1, I test the sensitivity of the results to an alternate definition of the instrument. I relax the restriction on surgeon age used in the definition and designate all surgeons who stop performing procedures (and who perform at least 5% of the hospital's CABG procedures) as exitors. This increases the number of exiting surgeons in the data to 63. The coefficient on surgeon volume is now smaller in magnitude, but still statistically significant (-.00025 vs. -.0007, p=.029).

In columns 2 and 3, I address the concern about using recent volume (as opposed to cumulative volume) to proxy for surgeon experience. In column 2, I use the procedure volume of the surgeon over the last *two* years and in column 3, I use the cumulative volume as measures of experience. In both cases I calculate the instrument in an analogous manner. Since I observe cumulative experience only for surgeons who start practicing after 1998, I restrict the sample accordingly in column 3. The coefficient on lagged surgeon procedure volume remains statistically significant (*p=.008*) and largely unchanged when I use the two-year measure instead of the prior year measure (-.00054 vs. -.0007). When I restrict the sample to surgeons who start practicing in 1998 or later (in column 3), the coefficient on lagged procedure volume has the right sign (-.00028 vs. -.0007) but is statistically insignificant (*p=.420*). This is probably an artifact of the smaller sample size in this specification.

Column 4 tests the sensitivity of the results to excluding all the control variables from the model. The coefficient magnitude is similar to that in the original specification (-.00062 vs. -.00070) but there is some loss in precision (*p=.098*). Finally, in column 5, I include surgeon-hospital interaction dummies in the regression instead of including them separately. The model is robust to this specification as well.

Table 2.6: Some Robustness Checks

Dependent Variable: Did the Patient die?

	(1) **Alt. Def.**	(2) **2 Year**	(3) **Cumul.**	(4) **No Control**	(5) **Surg-Hosp**
Surgeon Volume, prior year	-2.58E-04**	-5.48E-04***	-2.87E-04	-6.24E-04*	-7.15E-04***
	(1.18E-04)	(2.06E-04)	(3.55E-04)	(3.77E-04)	(2.70E-04)
Patient Characteristics	Y	Y	Y	N	Y
Hospital Characteristics	Y	Y	Y	N	Y
Surgeon Characteristics	N	N	N	N	N
Year Fixed Effects	Y	Y	Y	Y	Y
Hospital Fixed Effects	Y	Y	Y	Y	N
Surgeon Fixed Effects	Y	Y	Y	Y	N
Surgeon x Hospital dummies	N	N	N	N	Y
Number of Observations	107977	81739	13437	129184	107588

Note: All columns present results from IV regressions of patient mortality on total surgeon volume. Column 1 uses an alternate definition of exit. Columns 2 and 3 use volume from the last 2 years and cumulative volume instead of prior year volume as the measure of surgeon experience. Column 4 estimates the model with only fixed effects and no control variables. Column 5 estimates the model with surgeon-hospital fixed effects. Standard errors are reported in parentheses.
**** signifies p<.01, ** signifies p<.05 and * signifies p<.1*

I also tested the robustness of the results to the sample restrictions imposed earlier, by including low volume hospitals and surgeons in the sample and found that the main conclusions were not affected. In summary, the results are quite robust to alternate specifications. The only situation in which we lose statistical significance is when the sample is restricted to surgeons who start practicing in 1998 or later. In all other models, our main conclusions remain unchanged.

A possible alternate explanation behind the results is that exiting surgeons chose patients from zip codes whose inhabitants were healthier in unobserved ways. When these surgeons exit, the staying surgeons seem to experience an increase in quality (cause by the rise in procedure volume) when, in reality, they are now operating on patients who are healthier. I address this concern in two ways. First, the data show that patients of exiting surgeons had an average Charlson co-morbidity index of .894 compared to patients of staying surgeons who had an average Charlson index of .882. Since the patients of exiting surgeons seem to be as healthy as the patients of staying surgeons in observable ways, it is reasonable to assume that they do not differ significantly in unobservable ways. Second, the estimates show that increased procedure experience benefits surgeons across hospitals (and thus, across zip codes). This implies that the result is not driven by unobservable differences in patient health across zip codes.

2.7 Concluding Remarks

In this paper I examine the benefits of experience, one of the mechanisms through which individuals acquire (or maintain) competence in an organizational setting. Specifically, I study whether cardiac surgeons who perform more procedures experience an improvement in performance. In order to do so, I develop an instrumental variables estimation method that addresses the potential endogeneity of surgeon procedure volume. As my identification strategy, I consider exogenous shocks to the procedure volume of existing CABG surgeons in Florida caused by the exit of other surgeons from the same hospital. Using this instrument, I find evidence for a strong learning-by-doing effect for cardiac surgeons: an additional procedure a year leads to a reduction in patient mortality by 1.2%. I also find that procedure experience adds to a surgeon's general human capital, i.e. it has a positive influence on her performance across firms. In addition, I find evidence in support of some degree of task-specificity of surgeon human capital, i.e. a surgeon's experience performing one procedure has a positive effect on patient outcomes in other procedures, but this effect is smaller than the effect of experience performing the procedure in question.

These findings have implications for managers within the health care industry, specifically with respect to how to organize the firm so as to better leverage the skills of its professional workers. In highly specialized settings such as cardiac surgery, the knowledge retained by individual workers is often key to the successful functioning of the entire organization. While allocating procedure volume within a team, it is important for managers to bear in mind that the marginal benefit of additional experience is greater for low volume surgeons when compared to high volume surgeons. The quality of surgeons with low procedure volumes needs to be actively monitored. Hospitals can benefit from developing mentoring relationships between low-volume (or new) surgeons and more experienced surgeons. The more experienced surgeon can aid the low-volume surgeon in patient selection and also be available to assist in case of complications during the procedure.

The analysis presented in this paper has its share of limitations, one of the main ones being that the effect of hospital procedure volume on outcomes is not modeled. Since physician and hospital services are co-located, one cannot sort out their separate effects so easily. While some of the positive benefits of hospital volume on outcomes are captured by surgeon volume, others may not be. For example, larger hospitals may have a broader range of specialists and thus have experience with more diseases or co-morbidities and may learn to manage these better, leading to an improvement in patient outcomes. To the extent that such effects are not captured by hospital fixed effects and other hospital characteristics in our regression, our estimates of surgeon learning-by-doing will be biased. Since hospital volume affects outcomes in the same way as surgeon volume[38], we can think of the estimates in this study as representing an upper bound on surgeon learning-by-doing.

Another limitation pertains to the analysis of firm-specificity. The analysis assumes that the surgeon decision to operate across multiple hospitals is exogenous to her quality. This may not be true. One can imagine a situation in which surgeons who split their time across hospitals are of, say, higher quality (and are hence able to attract patients across hospitals). In that case, the estimates from the specifications in section 2.6.2 may be biased. A better strategy would be to instrument for the splitting decision. I leave this avenue open for further investigation.

A final limitation of the study is that CABG may be performed by teams consisting of two or more surgeons. Since the data only identifies the primary operating surgeon, I am unable to factor for the experience of other surgeons in the same surgical team who may have been involved in the procedure.

One of the questions that may arise about this study pertains to what exactly is being measured. Do the estimates reflect the core learning process of surgeons? Or do they represent the process by which surgeons maintain their skills, as opposed to learning new ones? The data does not allow me to disentangle these two effects. However, to the extent that surgeons perfect their procedural skills during (residency) training or the first couple of years of practice itself, most established surgeons

[38] Studies that have tested for the effects of both surgeon and hospital volume (in a correlational manner, without factoring for possible endogeneity biases) have found surgeon volume to be a much stronger predictor of patient outcomes

can be thought of as already being on the "flat" portion of their learning curves. Maintaining proficiency of skills by ensuring an adequate volume of procedures remains a recognized concern of surgeons. This concern arises due to the deterioration of skills with lack of practice: a surgeon who has an opportunity to consistently perform procedures over time will be able to maintain her abilities and skill level better than an equally trained surgeon who performs only a handful. Thus, the estimates may well measure the degree of "not-forgetting" as opposed to "learning".

This study adds to the large and growing theoretical and empirical literature that analyzes learning-by-doing in a variety of settings and industries. While the unit of analysis in most of these studies is the firm, this paper is able to look at learning-by-doing at the level of the individual worker, mainly due to the availability of detailed data. This study is to be viewed as a first step towards understanding the mechanism by which workers acquire and maintain their skills in organizations. While I have documented the existence of learning-by-doing in individuals, the link between individual and organizational learning is yet to be made. The data also lends itself well to exploring other aspects of learning in organizations, e.g. forgetting, learning spillovers etc.

Finally, one can use this setting to study whether the improvement in quality of the surgeon, caused by increased experience, translates into a greater share of patients in the market. More generally, one can quantify the extent to which learning-by-doing contributes to sustainable competitive advantage. I plan to explore these issues in future research.

3 The Substance of Style: A Study of Small Area Variations In The Practice Styles of Ob/Gyn Specialists[39]

3.1 Introduction

Since Wennberg and Gittelson (1973) demonstrated that a patient's likelihood of receiving a tonsillectomy varied from 7% to 70% across similar towns in Vermont, numerous studies have shown that the type of treatment a patient receives may depend on where the patient lives, and not just on what condition the patient has. (See Phelps and Mooney, 1993 for detailed reviews.) Much of this research considers variations across geographical locales (intra-state or inter-state) and finds that the rates of variation as measured by the coefficient of variation (CoV) range from less than .15 for well-understood procedures such as hip-fractures to CoV's higher than .50 for poorly understand procedures such as treatments for back-injury or diabetes (Phelps, 1999).

Many scholars have argued that such small area variations (SAVs) in practice styles[40] imply market failure and lead to welfare losses. For instance, Phelps and Parente (1990) estimated that the annual welfare loss in 1987 due to small area variations was $33 billion dollars and noted that this was an understatement if there is variation within a market and between markets. Health care economists have offered a plethora of solutions to remedy them ranging from standardizing medical curricula to

[39] Joint with David Dranove, Northwestern University and Hayagreeva Rao, Stanford University
[40] We use the term practice style to connote the propensity of a physician to treat a given patient in a particular way. For example, some physicians may have a propensity to deliver babies by cesarean section whereas others may favor vaginal delivery, all else equal.

altering incentives. The effectiveness of particular solutions would depend, naturally, on the reasons why SAVs exist, which is the subject of this analysis.

A number of traditional explanations for SAVs such as differences in patient preferences, income, underlying health status of physicians, physician density, access to medical care, availability of substitutes, or noise stemming from sampling have been disconfirmed in the literature (Phelps and Mooney, 1993; Wennberg and Gittelson, 1982; Bhikchandani, Chandra, Goldman, and Welch, 2001). Other writers have suggested that physician learning leads to convergence around community norms (Phelps and Mooney, 1999) but a recent study by Epstein and Nicholson (2005) shows that variation within a market is two to three times greater than variation across markets, and suggest that there is little revision of prior beliefs by physicians. Burke, Fournier and Prasad (2004) note that "while there is a vast literature describing the phenomenon, the puzzle itself largely remains unresolved".

SAVs are also an anomaly for sociologists, but have received scant attention (Flood and Fennel, 1997). Cultural-frame institutionalists suggest that social systems are likely to exhibit uniformity of structures and styles when individuals are subjected to either coercive pressure from regulators, normative pressures from professional networks, or mimetic pressure from peers (DiMaggio and Powell, 1983; Scott, 2001). The managed care model underlying the organization of medical care in the U.S. reflects a tight integration of regulative, normative and mimetic influences (Scott, Ruef, Coronna and Mendel, 2000). So the persistence of SAVs is an anomaly, and hence, an opportunity to study whether social diversity is based on spatial boundaries in a profession where there is a cultural pressure towards homogeneity. Medical sociologists have studied disparities in health care caused by race and status (Mirowsky, Ross and Reynolds, 2000), but a pressing challenge for medical sociologists is to understand "what might account for the large amount of unwarranted variation" in clinical practices (Shortell, 2004:14).

These considerations provide the motivation for our study. We begin by documenting the regional CoV in cesarean section rates across counties in Florida. We then assess whether SAVs are artifacts of individual-level variations among physicians, or artifacts of hospital-level variations, and address this issue using a simple but novel method that can be easily generalized to other procedures/regions. We specifically ask "Is the measured CoV when physicians and their patients are grouped by predetermined geographic region larger than the CoV obtained when physicians and hospitals are randomly assigned to "pseudo" regions that are equal in number and makeup to the actual regions?" We find that the median CoV in the actual data is roughly comparable to the median CoV's obtained through the random assignment of hospitals to pseudo-counties. This suggests that when it comes to c-sections in Florida, there do not appear to be any meaningful *regional* practice style effects; all variations in practice style are limited to variations across *hospitals*.

We then ask why are physician styles correlated by hospital. We focus on three major explanations. First, we test for matching of physicians to hospitals on the basis of style. Here again, we undertake a simple but novel approach that is easily generalized to other procedures/regions. We predict a physician's style on the basis of a vector of observable physician characteristics, including medical school and residency training identifiers. We correlate predicted styles of new physicians in year T, where T is the year in which they joined the hospital, with the actual styles in year T-1 of the hospitals that they joined, and find evidence of a substantial *matching effect*. (We think of this as a "doctors of a feather practice together" phenomenon.) We then test to discern if physician styles evolve over time – a *mimicry* or *learning* effect – and find them to be fairly stable. In particular, we find evidence in support of learning only in the short term, i.e. over the course of the first year a physician works at a hospital. This suggests that physician styles are imprinted early on in the careers of physicians and persist due to matching with the hospital rather than learning.

Finally, we find evidence that patients select physicians whose styles best match their own specific needs; i.e., a patient who is likely to require a cesarean will tend to select a physician whose style favors performing caesarians. This patient *selection* effect intensifies the measured SAVs that result from the physician matching described above.

3.2 Literature Review

The prevalence of SAVs in a wide range of medical specialties has occasioned a number of explanations. Early explanations hinged on variation in patient preferences. A study by Pritchard et al. (1998) refutes this argument. They asked patients at five medical centers their preferences for end-of-life care and found that the actual treatments were explained by the regional use of the form of care more than their preferences or clinical presentations. Another set of explanations implicate income and price variability across regions, but as Phelps (1999) notes, SAVs also exist in Canada where there is universal health care coverage. Yet another explanation is that SAVs are the outcome of random deviations from identical practices across communities, however, McPherson et al. (1981) reported that only 1-4 percent of the observed variation in Canada was due to noise.

Variations in illness patterns across regions are an intuitive explanation, but Fisher et al.'s (1994) study of a cohort of Medicare beneficiaries in the Boston area found wide variation across teaching hospitals, and little relationship between morbidity and actual hospitalization. Explanations hinging on access to medical care are also fragile because studies such as Bhikchandani et al. (2001) have shown that the CoV for myocardial infarction is lower than that for angina pectoris despite the fact both procedures require similar surgical and medical resources. Finally, variation in access to substitute procedures is another possible source of SAVs, but Phelps and Mooney (1993) show that the correlations among substitute procedures for ailments ranging from back-injury to strokes are positive rather than negative, thereby, implying that it is variation in physician perceptions rather than access to substitutes that matters.

Given the doubt cast on these traditional explanations, health care researchers have focused on the cultural foundations of spatial boundaries. Phelps and Mooney (1993) and Phelps (1999, 2003) postulate that physicians form beliefs about appropriate care during their medical and residency training, but learn from colleagues through Bayesian updating, and as a result, there is convergence around community norms. Since physicians with different training backgrounds will locate unevenly, there is likely to be inter-market variation in styles. Burke, Fournier and Prasad (2004) construct a formal model in which physician choices are shaped by a desire to conform to peers or spillovers of knowledge, and they show that small regional differences in patient mix give rise to different treatment patterns; so younger patients are less likely to receive bypass surgery when they live in an area where the average age of the population is higher.

There is some evidence for the fact that there is variability in resource use across medical schools in the U.S. Wennberg et al (2004) utilize claims data to document extensive variation across 77 Academic Medical Centers (AMCs) in the amount of care provided to Medicare FFS patients with three common conditions. They organize their analysis at the level of the hospital but acknowledge that practice styles of physicians, even those working at the same hospital, can differ and that this difference needs to be accounted for. If physicians develop their practice styles during medical school and residency training, variation across medical schools could act as a source of variation across physicians.

A recent study by Epstein and Nicholson (2005) of obstetrics and gynecology specialists in Florida shows that the variation in c-section rates across physicians within a market in Florida is two to three times greater than variation between markets. This suggests that physician-level variation may underlie regional variation. It also suggests that physicians do not substantially revise their beliefs due to local exchange of information, and as a result, physicians are unlikely to converge to a community standard, thus within-market variation is likely to persist. Epstein and Nicholson (2005) also report that residency training only explains four percent of the variation in risk-adjusted c-section rates across physicians, and note that although physicians learn from their peers, they do not significantly revise their beliefs about appropriate care.

Although Epstein and Nicholson's (2005) study is a valuable contribution, a number of challenges remain in understanding the social sources of SAVs. One set of issues concerns the level at which variation occurs; Epstein and Nicholson do not adequately consider a level of variation intermediate to the physician and the region, namely, the hospital. A corollary is whether SAVs at the regional level are a statistical artifact of variation at the hospital level.

A second set of issues pertains to whether the correlation among the styles of physicians in hospitals is simply an outcome of patient selection of physicians on the basis of unobservable patient characteristics (Burke, Fournier and Prasad, 2004), learning by physicians, or matching between physicians and hospitals. For instance, Epstein and Nicholson (2005) attempt to measure the extent to which physicians learn from their peers by regressing current practice style for each physician on the past practice styles of each physician's peers, and obtain a positive coefficient and infer that physicians are learning. Unfortunately, a positive coefficient is also consistent with selection and matching. We introduce methods that allow us to identify whether any or all of these phenomena are present.

Finally, we note that prior studies rely on hospital diagnostic coding to control for patient characteristics when measuring physician practice style. These studies neglect how variations in the coding practices and skills of hospital clerks may create hospital level correlations in "style" that belie the actual variations in physician practices. Thus, we rely on "transparent" patient characteristics that are unlikely to be subject to discretionary coding. While the use of transparent coding (as opposed to all available diagnostic information) does not affect our main results, failure to take this step with other data could introduce bias.

3.3 Overview of our Approach

Because we are offering a number of new analytic techniques, it is useful to lay out our overall approach before delving into details. We begin by presenting a statistical model that lays the foundation for the subsequent work. This model demonstrates how unobservable physician and patient characteristics can contribute to measures of SAVs. Our subsequent empirical analyses are designed to avoid the resulting ambiguities in interpretation.

Our first empirical task is to determine if measured CoVs are artifacts of physician or hospital level variations. We begin by computing the actual CoV for c-sections across counties in Florida. Our choice of county as the regional unit is purely for analytic convenience. We next compute the empirical distribution of CoVs when physicians and hospitals are randomly assigned to "pseudo" regions that are equal in number and makeup to the actual regions. If the actual CoV exceeds the 95^{th} percentile of the distribution of CoVs using pseudo-data, we conclude that there are genuine regional effects. Otherwise, we conclude that observed regional effects are artifacts of variation at a lower level of aggregation.

We then examine a series of possible explanations for CoVs. The most important of these is physician matching based on prior preferences. We could look at the correlation of physician styles within hospitals, but we would not be certain whether a positive correlation reflected matching based on prior preferences or other factors such as learning or the desire of physicians to conform to norms. To rule out the latter two explanations, we estimate a model predicting a physician's style on the basis of a vector of observable physician characteristics, including medical school and residency training identifiers. We then correlate predicted styles of new physicians in year T, where T is the year in which they joined the hospital, with the actual styles in year T-1 of the hospitals that they joined. We then examine whether the correlation increases over time to assess the extent of conforming/learning.

We conclude our empirical analysis by looking for evidence of patient sorting. We ask whether patients whose observable characteristics would indicate a preference for c-sections (e.g., they may have had a prior c-section) seek out physicians whose practice styles favor performing c-sections.[41]

We select Florida for our study of SAVs because the state makes available data that permits us to identify the physicians who perform surgical procedures. The Florida data is provided by the State Agency for Health Care Administration (AHCA). This data is comparable to patient level discharge data provided by the Health Care Utilization Project (HCUP) and the California Office of Statewide Health Planning and Development (OSHPD). Specifically, AHCA provides information about each discharge, including the hospital, diagnostic information, and demographic information (including the residence zip code.) In addition, AHCA provides the license number of the operating physician. We were able to obtain information about each physician's training and draw inferences about the

[41] In a clever study, Epstein and Nicholson (2005a) use information on practice variations in weekdays versus weekends as evidence of patient sorting.

physician's sex by linking the license to a separate online data base provided by the state of Florida. To protect confidentiality, we do not report any physician-specific information.

We study deliveries because there are many observations and there is a clear dichotomous decision – vaginal versus cesarean section – that may be used to identify practice "styles." A number of previous studies of practice variations have also examined the vaginal/cesarean dichotomy. We use data from 1994-2003 for various aspects of our study, but for convenience we document the extent of variations using only data from 2000-2001.

We consider each of the 67 counties in Florida to be a distinct market area. Previous studies of SAVs have used the county as the geographic unit. Using the county also allows us to focus on measuring the sources of variation, without dwelling on issues of market definition. As the main purpose is to suggest new analytic tools that can be applied to any procedure and market, we will not comment further on the appropriateness of using the county as the relevant geographic unit.

3.4 A Statistical Model

The typical analysis of SAVs examines both raw and regression-adjusted procedure rates. We can represent this analysis using a standard binomial choice framework. We suppose that the decision to have a medical procedure is the result of a decision making process that incorporates both patient characteristics (e.g., the patient's clinical condition) and physician characteristics (e.g., the physician's training and beliefs). Let the benefits of a procedure *B* (relative to an alternative intervention) be represented by:

(1) $$B = B_0 + B_1X_p + B_2X_d + \varepsilon_p + \varepsilon_d$$

where X_p are observable (to the researcher) patient characteristics, X_d are observable physician characteristics, and ε_p and ε_d are unobservable patient and physician characteristics respectively. The parameters B_1 and B_2 measure the importance of observable physician and patient characteristics.

The patient undergoes the procedure if *B* exceed some cost *C* (that may include financial cost as well as inconvenience, etc.) Thus, the patient has the procedure if

(2) $\varepsilon_p + \varepsilon_d > C - (B_0 + B_1X_p + B_2X_d)$

This can be estimated using a logit or linear probability regression model:

(3) $Y = f(X_p. X_d)$

where the dependent variable Y = 0 if the patient has a vaginal delivery and Y = 1 if the patient has a cesarean.

In practice, analysts do not include physician characteristics in their estimate of equation (3). Instead, the typical regression model takes the form:

(4) $Y = f(X_p)$

The magnitude of SAVs is computed from estimates of equation (4) roughly as follows. Let e_i represent the regression error for patient i. Let R_{ej} represent the average value of e_i for all patients residing in a region j. (Alternatively, one can include region fixed effects in the regression and use the fixed effects coefficients to measure the average value for each region.) Let M equal the mean intervention rate in the population. Then the coefficient of variation (CoV) of regression-adjusted practice styles is given by the standard deviation of R_{ej} divided by M. (Alternatively, it is the standard deviation of the county fixed effects divided by M.) The CoV is the measure of SAV used by most researchers.

The CoVs that result from estimation of equation (4) necessarily incorporate variation in all physician characteristics as well as variation in unobservable patient characteristics. Thus, if $\varepsilon_p = 0$ (i.e., there are no unobservable patient characteristics that influence the value of a cesarean), then the measured SAVs will reflect only variations in physician characteristics. Now suppose $\varepsilon_p \neq 0$. Measured SAVs will reflect variations in physician characteristics *as well as variations in unobserved patient characteristics*. If ε_p is uncorrelated with X_d and ε_d, then the presence of unobserved patient characteristics will simply add noise to the measured practice variations and no harm is done by ignoring it. If ε_p is positively correlated with X_d and ε_d, then this correlation will intensify the measured extent of practice variations. Such a positive correlation would occur if patients with unobserved preferences for caesarians seek out physicians who themselves prefer to perform caesarians. This is certainly a plausible supposition.

To summarize, most studies of SAVs control for observable patient characteristics. Thus, the reported CoVs reflect intra-region variation in all physician characteristics and intra-region variation in correlated but unobservable patient characteristics.

3.5 Preliminary Analysis

In this section we document the extent of practice variations for c-sections in Florida during 2000 and 2001.[42] We begin by computing the county-level coefficient of variation (CoV) of "raw" practice styles. To do so, we simply find the mean and standard deviation of raw cesarean section rates in each county.

[42] Throughout this analysis, we restrict attention to physicians who perform at least 20 deliveries in the two years combined. See footnote 11 for further explanation.

Variations in raw cesarean section rates might be attributable to differences in patient characteristics across market areas. To account for this, we estimate equation (4) using a linear probability model. To compute the "fully adjusted" CoV, we include among the predictors both patient demographics (age, income, race, insurance status and a vector of county dummies) and numerous clinical indicators (including whether the mother had a previous cesarean and whether hypertension was listed as a secondary diagnosis). We also include the medical liability insurance rate for that county as a regressor, in order to capture the effect of malpractice pressures on cesarean section rates.[43] To compute the *county-level* CoV of fully adjusted style, we compute the CoV of the county fixed effects.[44]

The coding of some of the clinical indicators used to adjust practice style may be subjective. If so, the reporting of these indicators could vary by hospital for reasons that had nothing to do with the patient's health or physician's style. Referring back to our model, this would imply that X_p is measured with noise that is correlated among physicians within a given hospital. This would introduce correlation in the measured practice styles of physicians within a hospital and could bias upwards the measured CoV. Thus, we also compute a "partially adjusted" CoV in which we restrict the predictors to demographic variables and a handful of unambiguous clinical conditions (specifically, previous cesarean and multiple gestation) that we believe are likely to be consistently coded across hospitals.[45]

Table 3.1: CoVs for Cesarean Sections in Florida

Risk –Adjusters Used	CoV
None	.189
Partial	.147
Full	.127

[43] We obtain data on malpractice insurance premiums for counties in Florida from the Medical Liability Monitor Annual Rate Survey. This survey divides Florida into three regions.
[44] We specifically compute each individual county's cesarean section rate at the mean of all predictors and compute the resulting CoV.
[45] We spoke with a "coding consultant" who works with hospitals to improve the accuracy of their coding and she confirmed that these codes are likely to be recorded consistently.

Table 3.1 reports the county-level CoVs of the unadjusted, partially adjusted, and fully-adjusted practice styles. We obtain a county level CoV of "raw" practice styles of 0.189, which is in line with previous estimates in the literature. The partially adjusted and fully adjusted CoVs are lower than the "raw" CoV, since we control for differences in patient characteristics across counties while computing these estimates. The fully adjusted CoVs are lower still, suggesting that controlling for all patient characteristics reduces observed variability, despite the potential for bias mentioned above. Even so, we will work with the partially adjusted CoVs in future analyses, preferring noise to bias. None of our key findings are materially affected by this choice.

3.6 Are SAVs Statistical Artifacts?

Physicians have their own individual practice styles. [46] It follows that there would be measurable (though possibly small) CoVs in cesarean rates across counties even if physicians were randomly allocated across the counties. By the same token, inter-hospital variation in practice styles will appear as county-level variations, even if hospitals are randomly located. It follows that an important step in assessing the magnitude of practice variations is to determine the level of aggregation at which variations occur. The levels of aggregation we consider, ordered from lowest to highest, are as follows[47]:

· *Individual variation*: Measured SAVs at the hospital or county level are an artifact of differences among individual physicians.

---

[46] Phelps et al. (1994) document enormous variation in practice "styles" from one physician to the next.

[47] We would like to also consider aggregation at the level of the group practice, but appropriate data are not available.

· *Hospital*: Physicians with similar styles practice at the same hospital. Measured SAVs at the county level are an artifact of differences among hospitals.

· *County*: Measured SAVs at the county level cannot be explained by differences at the hospital level. Styles across hospitals within a county are positively correlated.

County-level SAVs could be statistical artifacts of lower level variation whenever: (1) there are large differences in physician/hospital practice styles and (2) few physicians/hospitals account for a large percentage of the total procedures in some regions. If (1) and (2) hold, the large variations at the physician/hospital level will show up as county-level variation, even though there are not county-level effects. In general, there is no way to be certain *a priori* whether observed SAVs at the county level are artifacts of lower level variation. One must examine the data.

We offer a novel method for identifying the proper level of aggregation. To illustrate our method, consider a hypothetical situation in which there are substantial differences in practice styles across physicians and only a handful of physicians in each region. Specifically, suppose that half of all physicians perform an invasive procedure on 40 percent of their patients, all else equal. The other half performs the procedure on only 20 percent of their patients. If each region has four physicians, then we will observe the following distribution of procedure rates:

Rate of Intervention	**Distribution of Regions**
40%	6.25%
35%	25%
30%	37.5%
25%	25%
20%	6.25%

The coefficient of variation (CoV) for this data is 0.17. This is in line with the estimated CoV for caesarians in Florida as well as many other published estimates, but in this example it is entirely due to random locations of individual physicians; there are no regional effects.

Here is another way to think about it. The "region" is a potentially arbitrary way of grouping physicians and their patients. If region really does not matter, then grouping into regions should be no different, in a statistical sense, than any other arbitrary grouping. To determine if region really does matter, we ask the following question:

> Is the CoV when physicians and their patients are grouped by predetermined geographic region larger than the CoV when physicians are randomly assigned to "pseudo" regions that are equal in number and makeup to the actual regions?[48]

Random Assignment of Physicians to Pseudo-Counties

We treat each county as a region and reference the county level CoVs as reported in Table 3.1. Using the same data, we then compute the CoV for randomly constructed "pseudo counties" – random collections of physicians and their patients that resemble the actual counties in terms of number of physicians and patients. In this exercise, we use data aggregated over 2000 and 2001, and restrict our attention to physicians who perform at least 20 procedures in the two years combined.[49] We construct pseudo counties as follows:

---

[48] Note that in order to perform this exercise, we must "attach" patients to the physicians who treat them. Thus, the CoV in pseudo regions reflects random variations in both physician characteristics and unmeasured patient characteristics.

[49] We do this for analytic convenience. These physicians account for over 98 percent of all deliveries. The few remaining physicians would not have a dramatic impact on the measured CoV.

1) We assign physicians to volume "tiers" based on the number of procedures performed by the physician in a county[50]: 400+ procedures; 200-400 procedures; 100-200 procedures; 20-99 procedures.

2) We identify where each physician practices. Let N_{tc} represent the number of physicians in tier t who perform deliveries in county c.

3) We select the first county and for each tier t in that county, we randomly select N_{t1} of the physicians assigned to tier t. We assign these physicians to the first "pseudo county."

4) We repeat the exercise for each remaining county, sampling without replacement. Once we have done this for all 67 counties, every physician is assigned to a pseudo county.

In this way, we create 67 pseudo counties that correspond to the original 67 Florida counties in the sense that they have the same number of physicians in each volume tier. We then compute the CoV for the pseudo counties. We repeat this experiment 1000 times to obtain the distribution of CoVs that would be obtained if all variation was due to random locations of physicians. We compare the actual CoV to this distribution to assess whether the actual value is larger than what we would expect to observe due to random chance.

Table 3.2 reports the CoVs for the actual counties and summary statistics on the distribution of CoVs in the pseudo counties, based on 1000 repetitions of the method described above. Table 3.3 reports the same data, restricted to rural counties.[51] There are two important takeaways from Tables 3.2 and 3.3. First, the CoV in the actual data exceeds the 99th percentile CoV in the pseudo data. We conclude that practice variation is not entirely a statistical artifact of differences across

[50] A physician operating in two counties is treated as two physicians. Around 4 percent of physicians in our data operate in two counties.
[51] Counties having 3 hospitals or fewer were designated as rural counties.

physicians. That said, the median CoV in the pseudo data is roughly 60 percent of the CoV in the actual data. This suggests that a substantial portion of observed SAVs in the actual data is due to random locations of physicians, rather than to any genuine regional effects.

Table 3.2: CoVs in Actual and Pseudo Counties – Physician Level Aggregation

	Actual counties	Pseudo counties 50^{th} percentile	Pseudo counties 95^{th} percentile	Pseudo counties 99^{th} percentile
Raw Data	.189	.108	.146	.171
Transparent covariates	.147	.090	.111	.127
Full covariates	.127	.072	.097	.103

Table 3.3: CoVs in Actual and Pseudo Rural Counties – Physician Level Aggregation

	Actual counties	Pseudo counties 50^{th} percentile	Pseudo counties 95^{th} percentile	Pseudo counties 99^{th} percentile
Raw Data	.220	.127	.169	.191
Transparent covariates	.169	.103	.125	.142
Full covariates	.143	.080	.102	.117

Random Assignment of Hospitals to Pseudo-Counties

Now consider an intermediate level of aggregation – the hospital. We will first document practice variations at the level of the hospital. We then ask whether the CoV when hospitals and their patients are grouped by county is larger than the CoV when hospitals are randomly assigned to pseudo counties that are equal in number and makeup to the actual counties.

To do this, we repeat the preceding exercise at the level of the hospital rather than the physician:

1) We classify hospitals into four volume tiers: <1500 procedures; 1500-3000 procedures, 3000-6000 procedures and >6000 procedures.

2) We identify hospitals in each tier in each county. Let H_{tc} represent the number of hospitals in tier t in county c.

3) We select the first county and for each tier t in that county, we randomly select H_{t1} of the hospitals from tier t. We assign these hospitals to the first pseudo county.

4) We repeat the exercise for each remaining county, sampling without replacement. Once we have done this for all 67 counties, every hospital is assigned to a pseudo county.

Tables 4 and 5 report our findings for all counties and rural counties respectively. Tables 4 and 5 show that the actual CoVs are only slightly larger than the median CoVs in the pseudo data; the actual CoVs are smaller than the 95th percentile CoVs. There does not appear to be any meaningful region practice style effects; all variations in practice style are limited to variations across hospitals.

Table 3.4: CoVs in Actual and Pseudo Counties – Hospital Level Aggregation

	Actual counties	Pseudo counties 50th percentile	Pseudo counties 95th percentile	Pseudo counties 99th percentile
Raw Data	.189	.166	.198	.209
Transparent covariates	.147	.125	.152	.165
Full covariates	.127	.118	.135	.144

Table 3.5: CoVs in Actual and Pseudo Rural Counties – Hospital Level Aggregation

	Actual counties	Pseudo counties 50th percentile	Pseudo counties 95th percentile	Pseudo counties 99th percentile
Raw Data	.220	.197	.226	.241
Transparent covariates	.169	.142	.176	.194
Full covariates	.143	.129	.148	.166

To summarize, there are measurable practice variations in the raw and adjusted county level data. These variations are partially explained by differences in practice styles at the level of the physician. The remainder of these variations is explained by differences in styles at the level of the hospital. There is no meaningful additional variation at the highest level of aggregation, the county.

3.7 Hospital-level Practice Styles

The preceding analysis is consistent with the idea that physicians practicing at a given hospital have positively correlated practice styles. This can occur for a variety of reasons:

(1) Physician matching: Physicians and hospitals choose one another based on common practice styles.

(2) Physician learning: Physicians who are new to a hospital learn or adopt the practice styles of their colleagues.

(3) Patient selection: There is a positive correlation in the unobservable characteristics of patients who choose that hospital.

In this section and the next, we offer direct and indirect evidence of each phenomenon. We begin by deriving physician styles by estimating equation (4). Recall that the resulting styles could reflect physician characteristics or unobservable characteristics of their patients. We then conduct a number of tests to confirm or rule out each of the reasons for hospital-level correlations in style. While we are able to affirm or dismiss each reason, we can not completely parse them out. That is, we cannot definitely state "X percent of hospital style is due to matching, Y percent to learning, and 100-X-Y percent to selection."

The estimation of equation (4) is neither novel nor central to our analysis so we will merely summarize our findings.[52] A patient with a previous cesarean has a much higher probability of delivering by cesarean in her current delivery. Older women are slightly more prone to having

[52] The full set of regression results are available from the authors upon request.

caesarians, as are black and Hispanic women (in comparison to white women). Patients living in higher income zip codes have a slightly smaller probability of having a cesarean. Increases in liability premiums are associated with slightly higher cesarean rates. The large sample size assured statistical significance of all the aforementioned predictors.

Consistent with our earlier analysis that revealed hospital-level practice styles, we find that the correlation between each physician's own adjusted style and the styles of all other physicians at the same hospital is .418, which is significant at $p<.001$.[53] (We computed this using transparent covariates so as to avoid inducing a spurious correlation attributable to common medical record keeping.) This confirms that physicians within a hospital have similar styles.

Evidence on Physician Matching

The matching hypothesis holds that physicians match with hospitals that share the same style. To search for evidence of matching, we examine the practice styles of physicians who are new to a hospital and the styles of that hospital in the year they join. We say that a physician is new to a hospital in a given year if the physician performed at least 25 procedures in that hospital in that year and had never performed more than 5 procedures in that hospital in any prior year for which we have data. We identified 767 new physicians using these criteria. Of these, 152 appear to be "brand new" physicians – new to a hospital and less than two years removed from completion of residency training.

We find that the correlation in practice style between new physicians and the existing physicians on the staff of their hospital is .356, while the correlation in practice style between "brand new"

[53] When computing correlations between physicians and their hospitals, we exclude from the hospital those patients treated by the physician in question. In addition, we include only those physicians who treat at least 25 patients in a given hospital/year, so as to exclude physicians whose practice styles are estimated with substantial noise.

physicians and existing physicians is .376. Both correlations are highly significant and consistent with matching.[54]

These correlations are also consistent with two alternative hypotheses: (a) physicians immediately adopt the styles of their colleagues upon joining a new hospital (what one might call *mimickry* or immediate learning), and (b) there are correlated but unobservable characteristics of the hospital's patients. To find additional evidence of matching, we develop a model that predicts each physician's practice style based on "predisposing characteristics" – demographic and medical training variables.[55] If we assume that physician selects a medical school and residency program prior to determining which hospital they wish to practice at, then a positive correlation between predicted physician styles and actual hospital styles would be definitive evidence of matching. We will call this "training-based" matching for simplicity.

The dependent variable for this model is the physician's style computed using transparent covariates as described above. The unit of observation in our regressions is a physician-hospital-year triad. For this analysis, we restrict our attention to physicians who perform more than 25 procedures at a hospital in any given year. This results in a sample containing 9180 observations with information on 1041 physicians across 10 years.[56]

Our predictors include a vector medical school and residency training identifiers (for those schools and programs that trained at least 5 physicians), a foreign medical school indicator, sex[57] and year of

---

[54] An alternative explanation for these findings is that new physicians immediately mimic the styles of their senior colleagues. We cannot fully rule out this alternative; however, the evidence on training that we present below further supports the matching hypothesis.

[55] Phelps (2003) and others have observed that training can also have an important impact through a variety of mechanisms.

[56] These are the observations for which all variables have non-missing values. We were unable to obtain information on training programs for a few physicians, and they were excluded from this analysis

[57] While we lacked data on the physician's sex, we were able to make informed guesses based on the physician's first name. We had categories for male, female, and indeterminate.

completion of medical school[58], and an interaction between the decade of graduation and the residency program indicator[59]. While constructing the interaction variables, we ensured that each "cell" was sufficiently populated in order to avoid perfect prediction.

The model also incorporates detailed information on each training program (residency and medical school). We include indicator variables for the location (i.e. state) of the training programs in order to capture regional effects on practice style[60]. We include the ranking of these training programs as a linear predictor.[61] Finally, we also include a dummy variable that indicates whether or not a training program is affiliated with a university. The model also includes year of training indicators (aggregated into ten year periods) to capture time-specific factors that may affect practice style.

The results of this regression appear in Table 3.6. The medical school and residency indicators are highly significant; where you train affects how you practice. Male physicians tend to perform more caesarians. Physicians who underwent training at medical schools in foreign countries do not have significantly different practice styles in comparison to their American counterparts. The other coefficients indicate that the characteristics of the residency programs seem to be strong predictors of practice styles, especially in comparison to characteristics of the medical school. Finally, unranked training programs do not seem to have a different practice style in comparison to ranked programs, on average.

[58] In alternate specifications, we tried including the number of years since graduation as a predictor, but dropped that variable in favor of year of completion for reasons of fit.
[59] This interaction was computed for those residency programs that train at least 20 physicians.
[60] We lumped together states that had less than 10 physicians.
[61] We used data from residentphysician.com to determine ranking of residency programs – based on the total amount of grants and awards received. We obtained rankings of medical schools from US News' "America's Best Graduate Schools 2006" survey. Programs that do not appear in the rankings are assigned the median rank but are also assigned a dummy variable. This allows complete flexibility for the average impact of unranked schools.

Table 3.6: Determinants of Physician practice style – Regression results

Predictor	**Coefficient**
Male	.0148***
	(.002)
Foreign Medical School	-.0175
	(.044)
Rank of Residency program	-.0012***
	(.00018)
Rank of Medical School	.0005
	(.0004)
Residency hospital affiliated with university	-.0112**
	(.0042)
Is the residency program ranked – indicator	-.0050
	(.0048)
Is the medical school ranked - indicator	.0029
	(.0076)
Residency program indicators	Yes
Medical school indicators	Yes
Residency program – state indicators	Yes
Medical school – state indicators	Yes
Adjusted R^2 statistic	***.1579***
Number of Observations	***9180***

Notes: *Regression also includes year indicators and indicators for year of graduation as predictors. Sample restricted to all physicians who perform at least 25 procedures at a hospital in a year between 1994 and 2003. Standard errors in parentheses.*

**** - p < .001, ** - p<.01 and * - p<.05*

We recovered from this regression each physician's predicted style. We correlate predicted styles of new (to the hospital) physicians in year T, where T is the year in which they joined the hospital, with the actual styles in year T-1 of the hospitals that they joined. This correlation is .177, which is significant at $p<.001$. The correlation for brand new physicians is even stronger: .211. It is not surprising that these correlations are smaller than those reported earlier, as they are derived indirectly from a model predicting styles. Even so, the significant correlations provide definitive evidence of training-based matching of physicians to hospitals.

Physician Learning/Mimicry

Hospital-level styles may also result from learning, as new physicians gradually adopt the styles of their peers. Epstein and Nicholson (2005) attempt to measure the extent of learning by regressing current practice style of each physician on the past practice styles of peers. They obtain a positive coefficient. While this is consistent with learning, it is also consistent with matching. In this section, we propose a simple alternative way to detect learning.

We again focus on physicians who are new to their hospital and begin by examining immediate learning, or mimicry. Recall that the correlation in the practice styles of physicians in their first year at a new hospital and the practice styles of their colleagues was .356 (and .376 for brand new physicians). If we restrict attention to just the first three months after arrival, as opposed to the first year, this correlation is only.255 (and .203 in the case of "brand new" physicians). This suggests that physicians do learn from (i.e., mimic) their colleagues over the course of their first year at a hospital. Learning/mimicry might occur even faster (i.e., in the first three months) but we cannot be certain.

We now turn to the question of whether new physicians continue to "learn" from their peers over time. We correlate the practice style of new physicians for the first four years after joining a hospital, restricting attention to those years in which the physician performs at least 25 procedures. Tables 7 and 8 report the correlations. The correlation *remains more or less constant over time.* Taken together with the previous results on first-quarter correlations, we conclude that there is evidence in support of mimickry/learning in the short term, but that practice styles tend to be fairly stable over longer periods of time.

Table 3.7: Correlation between Physician Style and style of other physicians in Hospital
Note: t=0 denotes year of entry into hospital

	t=0	t=1	t=2	t=3	t=4
Correlation Coefficient	.3564	.3601	.3307	.3506	.3979

Table 3.8: Correlation between Physician Style and style of other physicians in Hospital – Brand New Physicians
Note: t=0 denotes year of entry into hospital

	t=0	t=1	t=2	t=3	t=4
Correlation Coefficient	.3763	.5501	.3541	.4254	.3736

Patient Selection

As our theoretical model suggests, practice variations can result from correlations between physician characteristics and unobservable patient characteristics. We might expect a patient who has personal preferences for a cesarean section to select a physician who is predisposed to perform caesarians, thereby increasing the observed CoV. [62] We cannot, of course, detect whether unobservable patient characteristics are correlated with physician characteristics. However, it stands to reason that if there are correlations between observable patient characteristics and physician characteristics, then there will be correlations with the unobservable patient characteristics and physician characteristics as well. Thus, we can provide indirect evidence that selection on unobservables occurs, even if we cannot definitely measure its magnitude.

To test for this, we take our initial estimates of equation (4) predicting whether a given patient would have a cesarean. We use the results to predict the probability of a cesarean for each patient. This is a measure of the patient's observable preference for a cesarean. We find that the correlation between observable patient preferences and the practice styles of the physicians they select is .068, which is significant at $p<.001$. Note that the set of available patient observables is limited. Patients surely know much more than we do about their own preferences, so the extent of patient selection may be much higher than the correlation of .068 would suggest. In any event, we infer from the correlation between patient observables and physician styles that there is likely to also be a correlation between patient unobservables and physician styles. If so, this will increase the measured extent of SAVs.

[62] Alternatively, doctors with particular styles may seek patients with matching needs.

3.8 Discussion

Our study was motivated by calls to understand the large amount of variation in clinical practices (Burke, Fournier, and Prasad, 2004; Shortell, 2004:14). Our main goal is to introduce several new techniques for measuring the extent, level, and sources of practice variations. We illustrate these techniques by examining obstetrics services in Florida. We do not assert that any of our findings will generalize to other services or regions. However, all of our methods can be generalized.

When it comes to c-sections in Florida, there do not appear to be any meaningful regional practice style effects; all variations in practice style are limited to variations across hospitals. This finding suggests that spatial boundaries by themselves may not matter much unless they hinge on social boundaries – in the case of Florida, organizational boundaries among hospitals were the progenitors of clinical variations.

A second implication of our study is that organizational variations among hospitals result from physician sorting. We also find that physician styles do not evolve over time. This suggests that physician styles are imprinted early on their careers and persist due to matching with the hospital rather than learning. Our analyses also revealed that patients select physicians whose styles best match their own specific needs; i.e., a patient who is likely to require a cesarean will tend to select a physician whose style favors performing caesarians. This patient selection effect intensifies the measured SAVs that result from the physician matching described above.

Taken together, these results inform theory and policy. One unmistakable implication for theories of practice styles is the role of imprinting. Phelps (2003) and a number of other scholars have attested to the significance of training as an antecedent of practice styles. Our study suggests that training has a "lock-in effect" that is redolent of imprinting. The ethologist Lorenz (1935) defined imprinting as a stamping process and argued that it was different from learning in two respects. First, susceptibility to imprinting is confined to a very limited period, usually, early in an organism's life. Second, once developed, imprinting is irreversible and so early experiences have a lock-in effect on organisms. Our analyses suggest that early physician training at medical school and residency – the formative period of a physician's career leads to irreversibility in styles. In turn, the salience of imprinting suggests that models of physician styles ought not to overweight the learning from peers at the expense of early career influences.[63]

Although our study is situated in the healthcare domain, it also speaks to the literature on organizational sociology. In recent years, a number of neo-institutional sociologists have suggested that organizational homogeneity is the outcome of coercive pressures from regulators, the mimetic effects of peers, and normative influences of professions (DiMaggio and Powell, 1983; Scott, 2001). Clinical decision making entails regulators, peers and professional influences and is therefore, an interesting setting to understand the scope and magnitude of homogeneity. Out study shows that professional training leads to heterogeneity across hospitals than homogeneity and indicates that the effects of peer learning are weak and that heterogeneity exists despite the presence of regulators. So neo-institutional research in sociology needs to recognize that professional influences can spawn variety rather than homogeneity.

Finally, our results have important implications for health care policy. The study of practice variations is motivated by the view that they may be socially harmful. Once we examine the sources of variation, this normative conclusion is not so clear. A root cause of practice variations is physician training. It is not obvious whether it is socially desirable for all physicians to receive

---

[63] Imprinting thus means socialization at the inception of a career Physicians get socialized in the hospitals and practices they join but as our analysis that does not matter much. It is socialization in graduate school and residency programs that is decisive.

identical training. The implications for experimentation and innovation would have to be explored. Variation is intensified by physician matching. We can easily imagine that this matching is socially desirable, as it can facilitate peer evaluation and thereby promote quality within the organization. Patient sorting also intensifies variation. The fact that variations in obstetrics practice in Florida are a hospital-level phenomenon suggests that patients have choices of local provider practice styles; our evidence of patient sorting suggests that some patients avail themselves of their choices. This can hardly be a bad thing.

4 Does the Market Punish Aggressive Experts? Evidence from Caesarean Sections[64]

4.1 Introduction

There are many markets in which a worker simultaneously diagnoses a customer's needs and recommends a product or service to meet them. Customers have limited information on which to judge the merits of the recommendation and may, as a result, consent to excessively costly or unnecessary services. These are known as "credence good" markets and plumbers, auto mechanics, and lawyers are just a few of the workers who face the resulting potential for conflict of interest. This problem has received considerable attention in medicine, where physicians simultaneously diagnose their patients' conditions and recommend treatments. Physicians stand to prosper by recommending and performing costlier treatments than are indicated by an objective diagnosis. This is referred to in health economics as "supplier-induced demand."

The presence of credence goods poses a theoretical conundrum. What prevents agents from always recommending more costly services? There are some obvious constraints such as the face validity of the recommendation (such as if a patient admits to mild acid reflux and the physician recommends a heart transplant) and the potential for litigation (e.g., malpractice after unnecessary surgery is botched). Within health economics, McGuire and Pauly (1991) suggest that physicians are constrained by ethical considerations. Also writing about demand inducement in health economics,

[64] Joint with David Dranove, Northwestern University

Dranove (1988) suggests patients may avoid physicians who have a reputation for doing too many high cost procedures. In other words, simple market forces may limit the ability of sellers to exploit their informational advantage.

In this paper, we examine whether the market does, in fact, punish overzealous sellers. We focus on the market for deliveries. Following Phelps (2003), who equates a physician's predilection to perform a high cost procedure to that physician's "practice style", we measure the practice styles of obstetricians in several counties in Florida. Focusing on Orange County (home to Orlando), we then estimate a model of consumer choice, where one of the factors weighing on patient's choice of provider is that provider's practice style. We find that maternity patients prefer not to visit physicians with aggressive styles, *ceteris paribus*. The effect is most pronounced for high income patients and HMO patients, two segments of the market that might be very attractive to some obstetricians. We obtain largely similar results in several other smaller counties.

4.2 Theoretical Background

Our empirical work is motivated by two separate literatures, those on medical practice variations and credence goods. Wennberg (1972) was perhaps the first to document the fact that treatments received by patients depended heavily on where they lived and who provided their care, and not solely on their objective medical condition. Since this seminal work, there has been a considerable literature documenting the extent of these practice variations.[65]

There is considerable debate as to the sources of practice variations. Phelps and Mooney (1993) and Phelps (2003) postulate that physicians form beliefs about appropriate care during their medical and

[65] See Phelps (2003) for a discussion of this literature.

residency training, but learn from colleagues through Bayesian updating, and as a result, there is convergence around community norms.[66] Wennberg et al (2004) offer supporting evidence of variability in resource use across medical schools. Practice variations within local markets may also be a response to heterogeneity in patient preferences. Epstein and Nicholson (2006) provide evidence that obstetrics patients have preferences over practice styles and select their physicians accordingly. Thus, we might expect physicians to knowingly differentiate their styles as a way of establishing a profitable "competitive position" in the market.

The theoretical literature on credence goods also bears on our work. Health care is a quintessential credence good, and the health economics literature on the topic uses the term "demand inducement" to capture the idea that physician experts can exploit information asymmetries by prescribing care that is not in the best interest of their patients. Roemer (1961), Fuchs (1978) and many subsequent studies suggest that physicians do induce demand, for example by recommending highly remunerative surgical services. Gruber and Owings (1996) study demand inducement for caesarians. Caesarians are a candidate for inducement because they tend to be more remunerative than the alternative vaginal delivery. For example, Gruber and Owings report that physician charges for caesarians in 1989 were a third again higher than the charges for vaginal deliveries. Physicians may also prefer the convenience of scheduled caesarians, which also tend to take less time to perform. Consistent with notions of inducement, Gruber and Owings find that obstetricians in states experiencing relative increases in supply responded by performing relatively more caesarians.

There is similar evidence of "inducement" outside of medicine. For example, Harrington and Krynski (2002) find evidence that funeral home directors induce consumers to choose burial over

---

[66] Burke, Fournier and Prasad (2004) construct a formal model in which physician choices are shaped by a desire to conform to peers or spillovers of knowledge, but do not explain where peers form their own practice preferences.

cremation and Bartels, Fiebig, and van Soest (2006) find that plumbers tend to "overprescribe" repairs for homeowners.

A crucial question in the credence good/demand inducement literature is why sellers do not always recommend the most costly available treatment. Discussing physician demand inducement, McGuire and Pauly (1991) cite ethical constraints, namely that physicians factor patient utility into their own utility function, and suffer a utility loss from inducement. Physicians might also be constrained by malpractice concerns, though some authors have suggested that caesarians pose less of a malpractice risk than vaginal delivery. It is possible that similar ethical and legal considerations could limit inducement by funeral home directors, plumbers, and other experts. Dranove (1988) suggests that the market might limit inducement. For this to occur, individual consumers would need to learn about the reputations of individual sellers, even though they lack substantial personal experience. Hubbard (2002) presents evidence that that this is exactly what occurs in the market for vehicle emissions testing. He finds that car owners steer their business towards emissions testers who tend to be more generous in giving out passing grades.

If car owners seek out testers with favorable reputations, it is reasonable to suppose that maternity patients might do so as well. Not only are the stakes considerably higher, preferences for practice styles may be very strong. In particular, McCourt et al. (2007) perform an extensive review of the literature on women's preferences for mode of delivery and find that women overwhelmingly prefer spontaneous vaginal birth. It is also reasonable to suppose that maternity patients are capable of performing a successful search. Like car owners seeking out emissions testing services, women with strong birthing preferences would have ample time to gather information about physician practice styles. They are also likely to know many other women who have recently given birth and would probably be informed about the mode of delivery.

Thus, we posit the following tension in the market for delivery services: Obstetricians stand to increase their earnings by performing more cesarean deliveries. In doing so, however, they stand to drive away maternity patients (either all or a sizable segment) who prefer obstetricians with more conservative practice styles. Our aim is to test whether the latter occurs and therefore serves as a natural limit on inducement.

4.3 Data

We obtained patient level hospital data for the years 1994-2003 from Florida's Agency for Health Care Administration (AHCA). AHCA data are similar to other state inpatient data that have been widely used in health services research. The dataset contains moderately detailed information about every hospitalization, including the patient's diagnosis related group, some secondary diagnoses, limited demographic information, the type of insurance (e.g., HMO), and the patient's residence zip code. AHCA data also include the license number of the "operating physician"; in the case of childbirth, this is the physician who performs the delivery.

We match the physician license numbers to an online licensing data base that gives us background information about each licensed physician in Florida, including their medical school and place of residency training. We obtained rankings of medical schools from *US News and World Reports*' "America's Best Graduate Schools 2006" survey. We used data from residentphysician.com to determine ranking of residency programs based on the total amount of grants and awards received. Programs that do not appear in the rankings are assigned the median rank.[67]

[67] Some doctors in our data received their training several decades ago. We do not have access to older rankings. We did observe that most of the top ranked medical schools have longstanding reputations for quality; thus, we do

4.4 Methods

Our ultimate goal is to assess how individual physician practice styles affect patient choice of physician. This suggests a three-step procedure:

1) Develop a measure of patient preferences for cesarean sections

2) Measure physician practice styles

3) Incorporate that measure in a model of patient choice

Specifically, we will do the following. First, we will use data from 1994-1997 and 2000-2003 to estimate a linear probability model that predicts whether a patient delivers by cesarean section. We will take the coefficients on patient characteristics from this model to predict patient preferences for caesarians in 1998 (one of the omitted years in the first step.) Second, we compute each physician's "practice style" in 1998, calculated as the difference between the actual and expected number of caesarians (where the expectation is based on the patient observables.) Third, we estimate conditional logit models of patient choice of physician in 1999, where the key predictor is the physician's practice style (computed from the 1998 data.) We now provide more details on our methods.

Measuring Patient Preferences

We use data from the entire state of Florida for the years 1994-1997 and 2000-2003. The dependent variable is a 0/1 indicator for whether or not the patient received a cesarean section. In selecting the

---

not believe that the rankings that we use are very far removed from the rankings that might have been constructed in the past.

predictor variables, we must bear in mind that we will be using the results of this first stage regression to compute the expected number of caesarians to be performed by each physician. In turn, patients will compare this expectation against the actual number to estimate each physician's "style."

Following the prior literature on patient preferences for interventions such as cesarean sections, we consider a large set of predictors, including patient demographic characteristics (age, race, insurance status, and the average income of the patient's residence zip code), clinical characteristics (a vector of secondary diagnoses), and a vector of physician fixed effects. We call this the "full" set of predictors. Patients may be privy to a much smaller set of information about the physician's patients, limited only to demographic variables and perhaps one or two readily observed clinical characteristics such as whether the physician handles many multiple gestations. We call this the "limited" set of predictors.[68]

Measuring Physician Practice Styles

We use the regression results from the preceding analysis to predict patient preferences for the year 1998, the year that we excluded from the first stage. We then compute $PSection_i$, the number of cesarean sections each physician i would have been predicted to perform had that physician abided by patient characteristics. We also compute $Section_i$, the actual number of cesarean sections performed by physician i and compute $Style_i = (Section_i - PSection_i)/Section_i$. Our key predictor variable

[68] Physicians may also vary in the extent to which they report clinical characteristics. Limiting the analysis to these limited predictors eliminates any potential for bias that might result. We exclude "prior cesarean" from the list of predictors because this is endogenous to the other patient characteristics.

in the third stage, $Excess_i = \max(0,(Section_i - PSection_i)/Section_i)$, represents the extent to which physician *i* overperforms cesarean sections.[69]

In addition to avoiding physicians with high values of $Excess_i$, patients may also avoid physicians who underprovide caesarians. To account for this, we include a variable $Deviation_i = \text{abs}((Section_i - PSection_i)/Section_i)$. Note that by including both $Excess_i$ and $Deviation_i$ in the choice models that follow, we allow for asymmetric patient responses to physicians who overprovide and underprovide caesarians. To compute the effect of an aggressive style, we add the coefficients on $Excess_i$ and $Deviation_i$. We can also compute the effect of a "passive" style (underperforming caesarians) by simply examining the coefficient on $Deviation_i$.

Some physicians perform only a handful of deliveries in a given year and patients may not be able to form firm estimates of their styles. We limit our analysis to those physicians who performed at least 50 deliveries in 1998. These physicians account for over 90 percent of the deliveries in the market that we study. The remaining physicians represent the "outside good" in the conditional logit models that we describe below.

Estimating Patient Choice

Our plan was to estimate separate patient choice models for each of several markets in Florida. We chose our "markets" so as to facilitate the choice analysis. In particular, we treated each county as a

[69] Some patients may have unobservable characteristics that lead them to have a preference for caesarians. This suggests that the variable $Excess_i$ may overstate the "true" extent to which physician *i* overperforms cesarean sections. We do not think this is a problem for our study for two reasons. First, we believe that when patients assess physician practice styles, they are not privy to other patients' unobservable characteristics and therefore will form estimates of style that are similar to those that we compute. Thus, it is appropriate to use our measures of $Excess_i$ in the patient choice equation. Second, the coefficient on $Excess_i$ will be a lower bound on the "true" effect provided that the "true" style is an attenuation of $Excess_i$ of the form $\gamma \cdot Excess_i$, where $0<\gamma<1$.

candidate market.[70] We then limited our attention to those counties with the following characteristics:

(1) The county has a central city with a population of at least 50,000

(2) There are at least 20 and no more than 200 "high volume" providers in a given year

(3) At least 75% of the women residing in these counties delivered their children at hospitals in the county

Based on these criteria, we identified the following seven counties (and central cities): Alachua (Gainesville), Brevard (Melbourne), Escambia (Pensacola), Lee (Fort Myers), Leon (Tallahassee), Orange (Orlando), and Volusia (Daytona Beach). It was also important in some specifications to exploit geographic diversification and hospital fixed effects. Of our seven counties, only Orange County (home to Orlando) had more than 50 zip codes or more than 5 hospitals. Thus, we restrict most of our analysis to Orange County. Even so, Orange County alone has approximately 100 high volume obstetricians (nearly the total of the other counties combined); thus, this restriction does not hinder our ability to generate significant findings.

We estimate a conditional logit model of patient choice of physician. Our key predictors are the measures of physician practice style, $Excess_i$, and $Deviation_i$. In order to allow for different effects of physician practice style on different patient segments, we include interaction terms between patient characteristics (such as income, age[71] and whether the patient is insured by an HMO) and each of $Excess_i$, and $Deviation_i$. Our control variables include travel times from each patient's resident zip code

---

[70] Given the criteria that follow, the decision to use the county as the basis for market definition is innocuous. We exclude metropolitan areas that are larger than counties on the grounds of tractability.

[71] We include income and age as categorical variables consisting of 3 categories. The "High" category consists of observations above the 75th percentile in both variables, while the "Low" category consists of observations below the 25th percentile. Observations in the middle constitute the omitted category.

to each physician's "primary" hospital (defined as the hospital at which the physician performed the most deliveries), and interaction terms between patient characteristics and travel time.

We control for the average overall quality of each physician (and each hospital) in the following way. In one specification, we include a vector of physician fixed effects and interaction terms between patient characteristics and hospital fixed effects.[72] However, since $Excess_i$, and $Deviation_i$ are both perfectly collinear with the fixed effect for physician *i,* we retain only the style interaction terms in this specification. This allows us to determine whether aggressive practice styles are more or less attractive to certain patient segments, but does not allow us to determine whether aggressive practice styles are more or less attractive overall.

In a second specification, we retain the primary predictors of physician practice style, $Excess_i$, and $Deviation_i$, and include a vector of physician characteristics (e.g. ranking of medical school the physician trained at, year physician graduated from medical school) in place of the physician fixed effects. While these variables control for physician-related factors that could affect patient choice, they do not completely capture physician quality. To the extent that $Excess_i$, and $Deviation_i$ are correlated with unobserved aspects of quality, the coefficients on these variables will be biased.[73]

4.5 Results: Patient Preferences

Table 4.1 presents the results of our linear probability estimates of the mode of delivery. Column 1 presents results using only a "transparent" set of covariates, while Column 2 contains a full list of predictors. Both columns include a vector of year dummies in order to control for aggregate trends.

[72] Note that we cannot include hospital and physician fixed effects in the same regression since they are collinear.
[73] We found no consistent pattern of correlation between $Excess_i$, and $Deviation_i$ and observable quality (residency rankings), suggesting that there may be minimal correlation with unobservable quality.

We also include a vector of physician indicator variables that capture time-invariant differences across physicians in their propensity to perform cesarean sections (controlling for patient characteristics) .While this model is used mainly to retrieve patient preferences for cesarean sections, the coefficient estimates on individual predictors are also of interest. We present a brief discussion of coefficients estimated using the full list of patient characteristics.[74]

The estimates on the year dummies (not presented in Table 4.1) indicate an overall growth of cesarean sections over time from 1994 to 2003. The coefficients on the variables measuring age of the patient (introduced into the model as categories) imply that older patients have a greater chance of having a cesarean section, all else held equal. Patients aged 40 and above have a 22% higher probability of undergoing the procedure compared to those under 20 years of age. Hispanic patients have a slightly higher propensity (2%) for cesarean sections when compared to White (and Black) patients. Patients insured with HMOs, PPOs and Medicare all have higher propensities for undergoing caesarians when compared with Medicaid patients (the omitted category), although the magnitudes are quite small (<.5%), except in the case of Medicare patients (8%).

As expected, the clinical characteristics of the patient are strong predictors of the probability of a cesarean section. For example, a patient diagnosed as bearing multiple fetuses has a 22% higher probability of receiving a cesarean section compared to a patient bearing a single fetus. Finally, patients originating from high income zip codes have a lower probability of receiving a cesarean: a 1-standard deviation increase in per capita income (measured at the zip code level) reduces the probability of receiving a cesarean section by .7% for patients residing in that zip code.

[74] The coefficients on variables that are present in both models are similar in magnitude and sign

4.6 Results: Patient Choice Model

As mentioned earlier, we estimated the patient choice models only for deliveries that took place in Orange County so as to maintain computational feasibility. We estimated four patient choice models: these models include either full or transparent covariates, with and without physician fixed effects. We briefly summarize the findings for control variables before moving on to the main results of interest, which we will present in detail. In general, patients strongly prefer to visit a nearby provider. Wealthier patients seem to have less distaste for travel, as do older patients.

Our main interest is the preferences of patients for physicians with aggressive practice styles. Recall that we compute these preferences by taking the sum of the coefficients on $Excess_i$ and $Deviation_i$. In the models with physician fixed effects, we are able to determine whether patients with particular characteristics have stronger or weaker preferences for aggressive style than do other patients. We study the following characteristics: HMO (versus other insurance), and income. We select these characteristics because they represent segments that some physicians may target for their practice. In the models without fixed effects, we can compute the effect of an aggressive style on all patients as well as the effect on HMO and wealthy patients.

Table 4.2 summarizes our findings for the fixed effects regressions, giving the direction, significance, and magnitudes of the effect of aggressive style. We find that HMO patients as well as high income patients have a statistically significant distaste for aggressive physicians (this holds for the models with full and transparent covariates). In order to better interpret the magnitudes of these coefficients, we calculate the difference in market share between a physician with average style and one who performs 10 percent more caesarians than predicted (i.e. an aggressive physician).

Table 4.1: : Estimating Patient Preferences (Linear Probability Model)

Dependent Variable: Did the Patient undergo a cesarean section?

	Transparent Covariates	Full Covariates
Age: 20-30	0.071***	0.059***
Age: 30-40	0.167***	0.144***
Age >= 40	0.262***	0.218***
Black	0.017***	-0.002*
Hispanic	0.019***	0.018***
HMO	-0.0001	0.002*
PPO	-0.0002	0.005***
Medicare	0.105***	0.075***
Multiple gestation	0.327***	0.269***
Malposition		0.552***
Hypertension		0.046***
Herpes		0.038***
Polyhdramnios		0.156***
Oligohydramnios		0.111***
Hemorrhage		0.321***
Prolonged pregnancy		0.074***
Diabetes		0.212***
Fetopelvic disproportion		0.596***
Fetal distress		0.226***
Trauma to perineum and vulva		-0.344***
Zip code Income (Unit: $1000)	-0.001***	-0.001***
Year FE	Y	Y
Physician FE	Y	Y
R-Squared	0.03	0.244
Number of Observations	1304151	1304151

* p<0.10, ** p<0.05, *** p<0.01

These estimates are presented in Table 4.3. The distaste for aggressive physicians seems to be quite strong: depending on the set of patient covariates (full vs. transparent) used in the estimation, an aggressive physician (as defined above) stands to lose ~10% of his patients who are insured by an HMO and 8-9% of patients who are wealthy. These estimates are also strongly statistically significant.

Table 4.2: Conditional Logit models of patient choice of physician (Models including Physician Fixed Effects)

	Transparent Covariates	Full Covariates
Excess * High Income	-1.11***	-.928***
Excess * Low Income	-2.31***	-2.23***
Excess * HMO	-1.13***	-.9422***
Excess * High Age	-.945***	-1.06***
Excess * Low Age	-2.53***	-2.71***
Deviation * High Income	.148***	.088**
Deviation * Low Income	-.236***	-.079*
Deviation *HMO	0.034	-.173***
Deviation * High Age	-.006	-.139***
Deviation * Low Age	-.128***	0.002
Travel Time	-.212***	-.214***
Patient Income * Travel Time	.002***	0.002***
Patient Age * Travel Time	.001***	.002***
HMO * Travel Time	.034***	.032***
Hospital FE * (Age, Income)	Y	Y
Physician FE	Y	Y
Number of Observations	1139319	1139319

* p<0.10, ** p<0.05, *** p<0.01

Based on these figures, we can estimate how much a physician would have to gain by performing caesarians if they were to gain financially from their aggressive style. For example, consider a county, where a 10 percent increase in the cesarean rate translates into a 3 percent decrease in the total number of patients. A physician would have to earn at least 30 percent more from cesarean deliveries to justify a more aggressive style on purely financial grounds. (To the extent that physicians value their time, they would require a smaller fee differential.) Gruber et al. (1998) and Keeler (1996) report data suggesting that the fee differential is typically smaller than this amount.

Table 4.3: % Change in Market Share of Aggressive Physician (Models including Physician Fixed Effects)

Note: An aggressive physician is defined as one who performs 10% more caesarians than predicted. P-values in parentheses

	HMO Patients	**High Income Patients**
All Covariates	-10.06 *(.001)*	-7.67 *(.001)*
Transparent Covariates	-9.94 *(.001)*	-8.73 *(.001)*

Tables 4.4 and 4.5 summarize our findings for the regressions without physician fixed effects. We are now able to estimate the effect of having an aggressive style on all patients as well as specific patient segments. Once again, the coefficients imply a strong distaste for aggressive physicians among all patients, as well as the specific patient segments we study. A physician who performs 10% more caesarians than predicted experiences a loss in share of 20% across all patients.

Tables 4.6 and 4.7 summarize the effects on demand of a passive style. The pattern is less consistent than it was for aggressive style, and is not as precisely estimated. No segment of patients seems to have a distinct distaste for passive providers with some segments weakly preferring them. Given that passive providers stand to earn less money on each delivery they perform, this suggests that some providers are choosing their style on other than economic grounds.

Table 4.4: Conditional Logit models of patient choice of physician (Models without Physician Fixed Effects)

	Transparent Covariates	Full Covariates
Excess	-2.293***	-2.198***
Excess * High Income	0.128	0.258
Excess * Low Income	-1.176***	-1.041***
Excess * HMO	0.138	0.2629
Excess * High Age	0.32	0.0643
Excess * Low Age	-1.001***	-1.31***
Deviation	-.066**	-0.0378
Deviation * High Income	0.153***	.1286***
Deviation * Low Income	-0.227***	-.101**
Deviation *HMO	0.001	-.169***
Deviation * High Age	-0.033	-.122**
Deviation * Low Age	-0.126***	0.0001
Travel Time	-0.199***	-0.200***
Patient Income * Travel Time	0.002***	0.002***
Patient Age * Travel Time	0.002***	0.001***
HMO * Travel Time	0.032***	0.030***
Hospital FE	Y	Y
Physician Characteristics	Y	Y
Number of Observations	1002672	1002672

* p<0.10, ** p<0.05, *** p<0.01

Table 4.5: % Change in Market Share of Aggressive Physicians (Models without Physician Fixed Effects)

<u>Note</u>: *An aggressive physician is defined as one who performs 10% more caesarians than predicted. P-values in parentheses*

	All Patients	HMO Patients	High Income Patients
All Covariates	-19.22 *(.001)*	-18.48 *(.001)*	-16.16 *(.001)*
Transparent Covariates	-20.18 *(.001)*	-19.11 *(.001)*	-18.01 *(.001)*

Table 4.6: % Change in Market Share of Passive Physician (Models including Physician Fixed Effects)

<u>Note</u>: *A passive physician is defined as one who performs 10% fewer caesarians than predicted. P-values in parentheses*

	HMO Patients	High Income Patients
All Covariates	-1.63 *(0.001)*	0.84 *(0.05)*
Transparent Covariates	0.32 *(0.199)*	1.42 *(0.001)*

Table 4.7: % Change in Market Share of Passive Physician (Models without Physician Fixed Effects)

Note: A passive physician is defined as one who performs 10% fewer caesarians than predicted. P-values in parentheses

	All Patients	HMO Patients	High Income Patients
All Covariates	-0.35	-1.93	0.87
	(0.37)	*(0.001)*	*(0.069)*
Transparent Covariates	-0.63	-0.62	0.83
	(0.03)	*(0.045)*	*(0.017)*

4.7 Discussion

We also note that because many patients have strong idiosyncratic preferences, (both in their distaste for inducement and in their taste for unobservable physician characteristics), aggressive physicians do not lose all their patients. This may explain why physicians adopt a variety of practice styles. Physicians with "good styles" have more patients but earn less money per patient, while aggressive physicians have fewer patients but earn more money from each. Moreover, if a physician were to change practice styles so as to pursue a particular segment – e.g., the physician becomes conservative so as to attract HMO patients – the number of patients per physician in that segment would decline, potentially making such a switch unattractive.

In its current form, the analysis presented in this paper suffers from an identification problem. While we interpret our findings as reflecting consumer distaste for aggressive physicians, an alternative explanation could be that physicians are inducing demand for cesarean sections in the face of falling market share. In ongoing analyses, we are trying to rule out this alternate explanation of our results. One solution is to do an indirect test for inducement. We estimated physician-level regression models (that looked at within-physician changes) relating the fraction of deliveries performed via cesarean section to the total workload of the physician in the previous time period. We found a negative relationship, i.e. a physician who experiences a drop in workload tends to perform more cesareans the next time period. However, the coefficient was economically very small and would not be able to explain the rather large effects we see in this paper.

In a second approach, we are trying to construct an instrumental variable for a physician's practice style. We plan to use the style of the residency program where the physician trained, as an instrument for the actual style of the physician. To this end, we are collecting data from the HCUP NIS (Nationwide Inpatient Sample) and matching it to the AHCA data. We expect this phase to take another 2 weeks. Results from this specification will be included in the next version of the paper.

5 Paying a Premium on your Premium? Consolidation in the U.S. Health Insurance Industry[75]

5.1 Introduction

The private health insurance industry in the US has experienced much consolidation in the last decade. The managed care backlash of the late 1990s has been followed by a steady growth in consolidation in the private health insurance market with a number of health plans being acquired by their larger rivals. A few recent examples include Anthem's acquisition by Wellpoint (2004) and UnitedHealth's merger with Pacificare in 2005. A Goldman Sachs report lists 22 major acquisitions occurring between 1995 and 2004. There is also some evidence to show that local insurance markets are highly concentrated and are becoming increasingly more so. In a recent paper, Robinson (2004) documents the existence of two, possibly related, trends in the private health insurance industry: the increase in local market concentration, and the growth in insurer profits and premiums over the last few years.

No empirical research has shown a causal link between these trends. Using a large, private database of insurance contracts representing 10+ million covered lives, we establish that link. Our analysis uses panel data on local insurance markets from 1998-2005. We examine the effects of changes in local market concentration on changes in insurance premiums for fully-insured HMO plans. The

[75] Joint with Leemore Dafny, Northwestern University and Mark Duggan, University of Maryland

former is instrumented using the number of insurance carriers in a market. We are currently working on extending the set of instruments by using mergers between national insurance carriers as a proxy for concentration, the underlying logic being that consolidation of these carriers is less likely to be correlated with trends in insurance premiums in any particular market.

The estimates from market-level OLS as well as IV regressions provide strong support for the hypothesis that increased market concentration of insurers leads to higher prices. Depending on the specification and the instrument used, the IV results imply a 3-8% increase in health insurance premiums for a one standard deviation increase in the measure of market concentration. Coefficients estimated from OLS specifications are smaller in magnitude implying a downward bias while failing to control for endogeneity.

The results from this study have important implications for policy and regulation. It is essential to understand the reasons behind increasing prices faced by consumers in the health insurance industry; rising premiums might affect private health insurance take up and add to the rolls of the uninsured. It could also affect Government outlays since Medicaid and Medicare outsource care to the private sector for some enrollees. Finally, the results also imply that horizontal mergers in these markets demand greater scrutiny by antitrust authorities.

This study contributes to the existing literature on price-concentration effects by examining this relationship within an important industry: private health insurance. By employing instrumental variable techniques on a novel dataset, we are able to establish a causal link between local insurer concentration and premiums charged in a market. Also, our use of mergers[76] as an instrument for

[76] The current draft of this paper uses the number of carriers in a market as an instrument for concentration. Future versions will contain more details about using mergers to instrument for concentration.

local concentration represents, to the best of our knowledge, a useful addition to the literature on applied econometrics.

This paper proceeds in six sections. Section 5.2 provides a summary of the related economic research. Section 5.3 presents the data and discusses issues of sample definition. Section 5.4 outlines the empirical analyses and discusses results from various specifications. Section 5.5 concludes.

5.2 Background

The present study draws from two streams of research: studies of price-concentration relationships in the Empirical Industrial Organization literature, and studies of insurer concentration and consolidation in the Health Services literature. In this section, we present a brief review of each.

5.2.1 Price-Concentration Studies in Industrial Organization

There is a large body of literature in empirical economics that examines the relationship between local market structure and prices. Several of these studies have found a significant positive relationship between market concentration and price. This result has been found to hold in industries as diverse as Airlines (Borenstein and Rose, 1994; Morrison and Winston, 1990), Banking (Hannan, 1991; Hannan and Liang, 1992) and Grocery Retailing (Cotterill 1986). However, much of the earlier work in this area assumes market structure to be exogenous with respect to price. Bresnahan (1989) and Schmalensee (1989) point out that unobserved shocks to costs or demand could influence prices as well as the underlying market structure, leading to biased OLS results.

Recent price-concentration studies aim at resolving this issue of endogeneity. Evans, Froeb and Werden (1993) present a formal analysis of the underlying sources of endogeneity and demonstrate the extent of the bias in the OLS results by estimating fixed-effects OLS and IV price-concentration regressions on airline data. They find that the IV fixed-effects estimator exceeds the OLS fixed-effects estimate by about 150 percent.

Davis (2004) looks at the relationship between the local market structure and the prices charged to consumers in the US movie theater industry. As his measure of market structure, he uses vectors of counts of movie screens owned by own and rival screens within various distances. Looking at within-theater variation in pricing, he uses second and fourth lags of the market structure variables as instruments for their current levels and finds that ownership structure has a statistically significant but economically small effect on admission prices charged to consumers.

5.2.2 Studies of Insurer Concentration

The Health economics and Health services literatures have a few existing studies that look at the degree of concentration in local insurance markets and its relationship with prices. Robinson (2004) uses a database of state regulatory filings to study state-level market structure of commercial insurance carriers (HMOs and PPOs) over 2000-2003. He finds that the largest firm controls at least a third of the market in almost 40 states in 2002-03. The top 3 insurance firms typically dominate each market, and control over 50% of total enrolment in almost all states. Using a variety of other sources, Robinson also documents an increase in insurer revenues and profits over the same time period. There is, however, no attempt to establish a causal relationship between the two trends.

Wholey, Feldman and Christianson (1995) examine the effects of HMO market structure on HMO premiums from 1988 to 1991. Using the number of HMOs in a market as their measure of competition, they find that more competition reduces HMO premiums. Using Interstudy data on all HMOs operating in the US, the authors specify regressions relating premiums to market concentration using the HMO as the unit of observation. In other words, the dependent and independent variables in the specifications are estimated as weighted averages over all the employers served by the HMO in all markets. The results from their regressions indicate that a larger number of HMOs in a market is related to lower premiums on average.

Our study attempts to answer a question similar to the one described above, but differs from it in three significant ways. First, we have access to a rich dataset that provides us with detailed information on characteristics of the healthplan and of the underlying insured population; this enables us to include a rich set of control variables in the specifications. Second, we are able to carry out our analyses at the market level and are thus able to avoid any noise or errors involved in aggregating HMO data across markets. Finally, and perhaps most importantly, we address the issue of endogeneity of the market concentration measure by employing an instrumental variables approach.

5.3 Data

The primary dataset used in this study was provided by a major benefits consulting firm on a confidential basis. This is a panel dataset (spanning the years 1998-2005) containing information on the health insurance purchase decisions of a large set of employers in the United States, including a large portion of employers in the Fortune 500. The unit of observation in this data is a health plan, which represents a unique combination of an employer, geographic market, insurance carrier and plan type (HMO, PPO etc.). Dafny (2007) provides more details on the variables contained in this dataset.

5.3.1 Estimation Sample

We use the geographic market as the unit of analysis while studying the effects of insurer concentration on pricing of health plans. To that end, we use the plan-level dataset (described above) to construct a panel dataset where each observation represents a geographic market in a particular year. These geographic markets are defined by the benefits consulting firm that supplied the data and correspond to the geographies used by carriers and employers while negotiating rates. These markets most commonly correspond to US metropolitan areas. While concentration variables (e.g. HHI, 4 firm concentration ratio) are already measured at the level of the geographic market, variables measuring plan-specific information (e.g. premium, plan benefits) are calculated as market-level averages, where each plan is weighted by the number of employees enrolled in it.

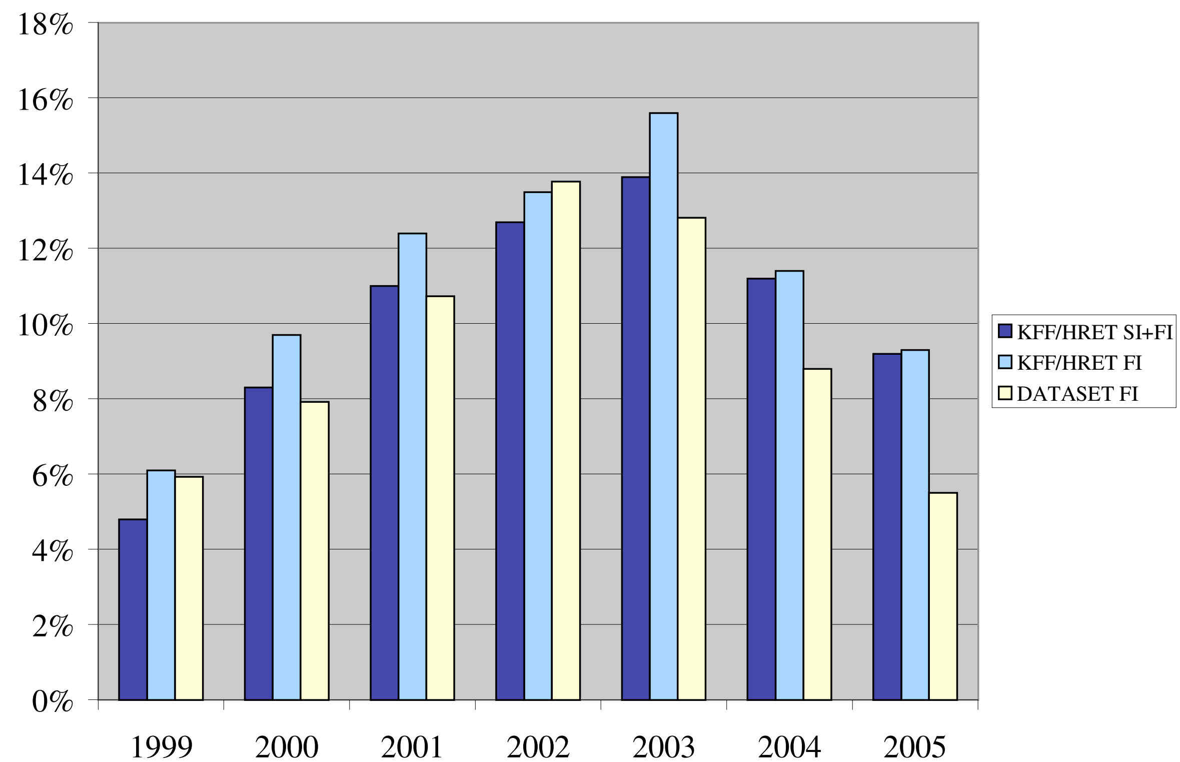
We restrict the estimation sample to fully insured HMO plans in geographic markets containing at least 20 employers offering one or more fully insured health plans. We study only fully insured plans because a true "premium" does not exist for self-insured plans where the employer pays the realized costs of managing the health of its employees. On the other hand, premiums are determined at the start of the calendar year for fully insured plans.[77] We impose the "20 client" restriction in order to ensure that our estimates of market structure are fairly accurate. Only 3% of the fully-insured enrollees end up getting dropped from our sample as a result of this restriction. Finally, we restrict attention to HMO plans for two main reasons: (a) Over 90% of the plans in our study sample are HMO plans and (b) In subsequent analyses, we are able to construct more accurate measures of concentration for the entire universe of HMO plans in the US using alternate data sources.

[77]. Refer Dafny (2007) for more details

5.3.2 Is the sample representative?

Since our dataset contains only a sample of the privately insured population, we compare some of the trends in our data to those reported by publicly available sources in order to ensure that our results are generalizable. The annual Employer Health Benefits Survey, sponsored jointly by the Kaiser Family Foundation (KFF) and the Health Research and Educational Trust (HRET)[78] contains some information on private healthplan premiums and benefits.

Figure 5.1. Growth in Annual Health Insurance Premiums, 1999-2005



<u>Notes</u>: KFF/HRET growth based on average premiums for a family of four, as reported by survey participants. "FI" denotes fully-insured plans, while "SI" denotes self-insured plans. "Premiums" for SI plans reflect employers' estimates of the cost of coverage. Dataset FI refers to fully insured plans in our Dataset. These figures are based on average premiums per covered employee, weighted to reflect the number of covered employees in each plan.

---

[78] Footnote on Kaiser survey

Figure 5.1 presents a comparison between the growth in premiums for privately insured plans calculated using our dataset, and as reported by the KFF-HRET survey. The premium growth in our data tracks the trends published by KFF/HRET fairly closely.

In addition, we undertake an exercise to see whether any geographical region in the United States is disproportionately represented in our dataset. We match the geographic markets in our data to counties[79] as defined by the Current Population Survey (CPS) conducted by the US Census Bureau. We then compare the number of enrollees in our data in each county to the total number of insured lives in that county. The estimates for county level insurance coverage are taken off the Small Area Health Insurance Estimates (SAHIE) conducted by the US Census Bureau which provides model-based estimates of health insurance coverage for counties and states.[80] We carry out this comparison for the year 2000 which is the only year for which the SAHIE estimates are available.

We compute a variable, ENROL_RATIO, which is calculated as the number of enrollees in our data in a particular geographical market divided by the total number of insured lives in that market.[81] The average county in our data has a value for ENROL_RATIO just over 1%. Since the SAHIE data contains estimates for the total number of insured lives, we believe our sample represents a greater portion of all privately insured lives in the US. When we charted the value of ENROL_RATIO across all counties in the US, we could not see any significant disparity across regions.[82]

[79] We matched markets to counties by comparing names. Using this procedure, we were able to match 2885 (out of a total of 3225) counties in the US to markets in our data. The unmatched observations represent counties which are not represented in the data – a good portion of these are counties in U.S territories like Puerto Rico, Guam etc.
[80] More information can be accessed at <http://www.census.gov/hhes/www/sahie/index.html>
[81] We arrive at this figure by summing the number of insured lives in all counties corresponding to that particular geographic market
[82] Chart to be included in later versions of this paper

5.4 Empirical Analyses

5.4.1 Some Trends in the Data

Before discussing the results, we examine some raw patterns in the data. The final estimation sample (Unit of observation: market-year) used for the analyses in Section 5.4 consists of 817 observations spread over 8 years. Table 5.1 provides summary statistics for some key variables in the sample. All plan-related variables are computed as enrollee-weighted market-level averages. The dependent variable in our analyses is the annual premium which is originally present in the dataset as a per-employee average, and includes both employer and employee contributions. The main control variables include *Plan Design,* which captures the average generosity of the benefits offered by plans in a market and *Demographic Factor,* which is a summary measure reflecting characteristics of the insured population (e.g. family size, gender). The decline in average *Plan Design* over time probably reflects benefit-cutting measures introduced by employers in order to contain costs.

We use two alternate measures of market concentration as our primary predictors: (a) the Herfindahl index (HHI) constructed using shares of insurance carriers in each market-year and (b) the 4-firm concentration ratio constructed using shares of the top 4 carriers in each market-year. By either measure, insurance markets in our sample experienced steady growth in consolidation over time. The average value of the HHI increased from .32 in 1998 to .39 in 2005, while the average four-firm concentration ratio grew from .90 in 1998 to .96 in 2005. Since the HHI and the 4 firm ratio are built off enrolment numbers for plans present in the data (as opposed to the entire universe of health plans), we also look at the number of carriers in each market to get a somewhat more accurate idea of the degree of concentration in each market. The average number of carriers in each market dropped from 8.5 to 6.4 over the same time period implying an increase in concentration over time.

Coincident with this increase in consolidation, the data also indicate that the average dollar premium (unadjusted) paid out by a firm in our sample increased from $3822 in 1998 to $6481 in 2005. This growth in premium over time persists even after we control for characteristics of the insured population, plan, market etc (Table 5.2)[83].

Table 5.1: Summary Statistics for Key Variables (Unit of Obs: Market-year)

(Sample: Fully Insured HMO plans in markets with 20+ distinct employers)

Year	HHI	4 Firm Concentration Ratio	Carrier Market Share	Premium ($)	Demographic Factor	Plan Design	Number of Carriers/Market	Number of Plans/Market	Number of Obs
1998	0.32	0.90	0.32	3821.76	2.31	1.13	8.48	77.76	108
	0.11	*0.08*	*0.11*	*421.68*	*0.17*	*0.01*	*2.86*	*49.64*	
1999	0.30	0.89	0.30	4058.82	2.27	1.13	9.14	87.21	117
	0.10	*0.08*	*0.10*	*467.48*	*0.13*	*0.01*	*3.19*	*53.44*	
2000	0.31	0.91	0.31	4207.07	2.23	1.12	8.01	75.31	109
	0.10	*0.07*	*0.10*	*555.85*	*0.13*	*0.01*	*2.74*	*43.62*	
2001	0.33	0.92	0.33	4474.01	2.25	1.13	8.12	84.29	113
	0.11	*0.07*	*0.11*	*575.83*	*0.12*	*0.01*	*2.79*	*55.91*	
2002	0.35	0.93	0.35	5091.55	2.28	1.13	7.67	82.19	110
	0.11	*0.06*	*0.11*	*681.98*	*0.14*	*0.01*	*2.47*	*56.79*	
2003	0.36	0.93	0.36	5621.05	2.28	1.11	7.46	77.41	101
	0.13	*0.07*	*0.13*	*738.06*	*0.13*	*0.02*	*2.61*	*56.11*	
2004	0.36	0.94	0.36	6291.13	2.36	1.12	7.12	60.02	83
	0.13	*0.06*	*0.13*	*1016.89*	*0.17*	*0.01*	*2.24*	*42.75*	
2005	0.39	0.96	0.39	6480.82	2.34	1.08	6.41	59.68	76
	0.14	*0.06*	*0.14*	*1093.67*	*0.16*	*0.02*	*2.14*	*43.31*	
Total	*0.34*	*0.92*	*0.34*	*4888.94*	*2.28*	*1.12*	*7.89*	*76.76*	*817*
	0.12	*0.07*	*0.12*	*1151.91*	*0.15*	*0.02*	*2.78*	*51.64*	

[83] In all calculations, we deflate the premium by the Consumer Price Index (Urban) to account for inflation.

Finally, as can be seen from the last column, the number of markets that fulfill the criteria specified earlier falls sharply towards the end of the study period. In Section 5.4.5, we test the robustness of our results to excluding these years from the sample.

Table 5.2: Premium growth over time (controlling for various factors)

Dependent Variable: ln premium

Year = 1999	0.060***	0.059***	0.083***	0.072***
Year = 2000	0.094***	0.120***	0.140***	0.144***
Year = 2001	0.155***	0.163***	0.193***	0.185***
Year = 2002	0.284***	0.296***	0.303***	0.305***
Year = 2003	0.383***	0.420***	0.403***	0.424***
Year = 2004	0.492***	0.523***	0.461***	0.502***
Year = 2005	0.520***	0.619***	0.505***	0.592***
Plan Design		1.841**	0.054	1.332***
Demo Factor			0.599***	0.419***
Market FE				Y
R-Squared	0.65	0.663	0.796	0.957
Number of Obs	817	817	817	817

* p<0.10, ** p<0.05, *** p<0.01

5.4.2 Relating Market Concentration to Price: OLS Estimates

We test for market power among insurers by estimating regression models that relate average premiums paid in a geographic market to the degree of concentration in that market. As mentioned earlier, plan-related variables such as *Premium, Plan Design* etc are measured as market-level averages where each plan is weighted by the number of enrollees. Equation (1.0) presents the basic OLS specification, where *m* indexes a market and *t* refers to a year.

We use the logarithm of the average premium paid in a market as our primary dependent variable. The key predictor is a market level measure of carrier concentration. While computing concentration measures, we make use of the entire set of fully insured HMO plans. However, we estimate our regression models on a restricted set of markets (restrictions defined in Section 5.3) because our concentration variables are measured with more accuracy for this subset. In order to account for the effect of plan generosity and characteristics of the insured population on premiums, we include *Plan Design* and *Demographic Factor* as regressors in all specifications. All models include fixed effects for each year to control for aggregate time trends, and for each market to control for local market conditions (e.g. wages) that might affect medical costs and hence premiums.

A positive sign on β_1 is consistent with the "market power" hypothesis. In Columns 1 and 2 of Table 5.3, the sign of the coefficient on market concentration is positive and statistically significant. In other words, an increase in concentration (as measured by an increase in HHI or 4 firm concentration ratio) is associated with a rise in premiums. A one standard deviation increase in HHI (4 firm concentration ratio) is associated with a .55% (1.15%) increase in premiums. Column 3 uses a non-linear measure of concentration by constructing quintiles of the 4 firm concentration ratio. Note that the coefficients in this model are identified off markets that switch quintiles because we include market fixed effects as predictors. The estimates reveal that markets that switch into the fifth concentration quintile (i.e. become more concentrated) have premiums that are 3.7% larger than markets in the least concentrated quintile.

Table 5.3: The effect of consolidation on premiums: OLS Estimates

Dependent Variable: ln premium

HHI	0.045*		
4 Firm Concentration		0.166***	
4 Firm Conc (quintile 2)			0.017**
4 Firm Conc (quintile 3)			0.017**
4 Firm Conc (quintile 4)			0.027***
4 Firm Conc (quintile 5)			0.037***
Year = 1999	0.073***	0.075***	0.075***
Year = 2000	0.143***	0.142***	0.144***
Year = 2001	0.184***	0.182***	0.184***
Year = 2002	0.304***	0.300***	0.302***
Year = 2003	0.421***	0.418***	0.419***
Year = 2004	0.499***	0.494***	0.496***
Year = 2005	0.587***	0.581***	0.581***
Plan Design	1.310***	1.315***	1.330***
Demo Factor	0.414***	0.415***	0.417***
Market FE	Y	Y	Y
R-Squared	0.957	0.958	0.958
Number of Observations	817	817	817

* p<0.10, ** p<0.05, *** p<0.01

5.4.3 Endogeneity Concerns

One concern over using concentration as a predictor in this regression is the potential endogeneity of that variable with respect to price. An unobserved shock to health plan costs could lead to a rise in health plan premiums which could force consumers to switch to cheaper carriers, leading to a change in concentration. This leads to a correlation between the concentration measure and unobserved determinants of premiums, leading to biased OLS estimates. A second source of endogeneity has to do with the fact that insurance carriers can enter and exit the market at any given time, and these entry and exit decisions could well be correlated with expected premium growth (or decline) in these markets.

An ideal instrument should explain variation in market concentration, while being unrelated to any unobserved determinants of price changes. We posit the use of the number of carriers in a market as a potential instrument for concentration. We believe that this variable addresses the endogeneity issue to some extent, because a price shock would need to induce entry (or exit) in order for the instrument to be correlated with the error term. We also use categories of the number of carriers in a market (<6,6-10,>10) as an instrument for concentration.

It should be noted that any bias that results from induced changes in concentration works to make the "market power" effect seem weaker than it is in reality. In other words, a positive price shock might lead to entry of new carriers resulting in a drop in market concentration. This would then introduce a downward bias on the coefficient measuring the impact of market concentration on price. Our coefficient estimates can thus be taken as a lower bound on the effect of market concentration on health plan pricing.

5.4.4 Relating Market Concentration to Price: Instrumental Variable Estimates

Table 5.4 presents results from first-stage regressions of market concentration (HHI and 4 firm concentration ratio) on our instruments, along with other exogenous predictors from the second stage. Columns 1, 2 and 3 present results from specifications that use HHI as the dependent variable and the number of carriers in the market, introduced respectively as a continuous variable, with a squared term and as categories, as predictors. The coefficient on the number of carriers is of the expected sign (negative, implying that markets with fewer carriers have higher concentration levels) and strongly statistically significant. The F-statistic for instrument significance is significantly larger than the typical recommended thresholds. Columns 3 and 4 contain results from specifications that use the 4 firm concentration ratio as the dependent variable. The coefficients are similar in magnitude to the ones estimated with HHI as the dependent variable.

Table 5.4: First Stage Regression Estimates of Market Concentration on Instruments

	Dep Var: HHI			Dep Var: 4 Firm Conc		
Number of Carriers	-0.015***	-0.049***		-0.015***	-0.024***	
Number of Carriers Squared		0.002***			0.0005***	
Number of Carriers category (6-10)			-0.044***			-0.037***
Number of Carriers category (>10)			-0.062***			-0.064***
Plan Design	0.689**	0.665**	0.646*	0.303	0.296	0.275
Demo Factor	0.127***	0.141***	0.116***	0.038*	0.042**	0.029
Year FE	Y	Y	Y	Y	Y	Y
Market FE	Y	Y	Y	Y	Y	Y
R-squared	0.667	0.683	0.655	0.72	0.723	0.684
Number of Observations	817	817	817	817	817	817

* p<0.10, ** p<0.05, *** p<0.01

The complete results from the IV regressions are presented in Table 5.5. Columns 1 and 2 contain coefficient estimates where market concentration (measured using HHI in column 1 and 4 firm concentration ratio in column 2) is instrumented for by the number of carriers in a market. The estimates provide strong support for the market power hypothesis: an increase in HHI (4 firm concentration ratio) of one standard deviation leads to a 3.9% (2.3%) increase in premiums. Since we include market fixed effects in the specification, this coefficient reflects within-market changes. The next two columns add a squared term (number of carriers squared) to the list of instruments. The coefficient on 4 firm concentration ratio is relatively unchanged but the coefficient on HHI drops slightly. The last two columns of the table introduce the instrument (number of carriers) as a categorical variable. The coefficients are similar in sign and magnitude to the estimates obtained by introducing the instrument as a continuous variable. Comparing these results with the OLS results, it is clear that failing to account for endogeneity introduces a substantial downward bias in the magnitudes of the coefficients.

Table 5.5: The Effect of Insurer Concentration on Premiums: IV Estimates

Dependent Variable: ln Premium

	IV = Number of Carriers		**IV = (Number of Carriers, Number of Carriers Squared)**		**IV = Number of Carriers (Categories)**	
HHI	0.324***		0.247***		0.338**	
4 Firm Conc		0.331***		0.335***		0.354***
Year = 1999	0.077***	0.077***	0.076***	0.077***	0.078***	0.077***
Year = 2000	0.140***	0.141***	0.141***	0.141***	0.140***	0.141***
Year = 2001	0.179***	0.179***	0.181***	0.179***	0.179***	0.178***
Year = 2002	0.295***	0.296***	0.297***	0.296***	0.294***	0.295***
Year = 2003	0.404***	0.413***	0.409***	0.413***	0.403***	0.412***
Year = 2004	0.482***	0.487***	0.486***	0.487***	0.481***	0.486***
Year = 2005	0.557***	0.569***	0.565***	0.569***	0.555***	0.568***
Plan Design	1.176***	1.299***	1.213***	1.298***	1.169***	1.296***
Demo Factor	0.383***	0.411***	0.391***	0.411***	0.381***	0.411***
Market FE	Y	Y	Y	Y	Y	Y
Number of Obs	817	817	817	817	817	817

* p<0.10, ** p<0.05, *** p<0.01

5.4.5 Robustness Checks

We test the robustness of the estimates through some sample restrictions. First, WE restrict the sample to years earlier than 2004. This is because the number of markets in our data experiences a significant drop in that year and stays that way for 2005. The first 2 columns of Table 5.6 contain OLS and IV results for this sample restriction. We focus on the coefficient of interest and use HHI as the measure for market concentration. The instrument used is the number of carriers in a market. As can be seen, the results are fairly similar to the ones reported in Tables 3 and 5, indicating that the analysis is robust to this sample exclusion.

Table 5.6: Some Robustness Checks

Dependent Variable: log(premium)

	Year < 2004		Markets with 20+ FI/SI Plans	
	OLS	IV	OLS	IV
HHI	.052**	.249**	.049**	.194**
Year FE	Y	Y	Y	Y
Market FE	Y	Y	Y	Y
Plan Characteristics	Y	Y	Y	Y
Number of Obs	658	658	1077	1077

Note: Number of carriers in the market is used as an instrument for HHI

* p<0.10, ** p<0.05, *** p<0.01

Finally, in the last two columns of Table 5.6, we use a slightly larger estimation sample: we now look at fully insured HMO plans in market-years that have at least 20 employers offering fully insured or self-insured plans. The concentration measure (HHI) is now computed using the set of fully insured and self-insured plans, and we now have 1077 market-years that fulfill the above criterion. Note that we restrict ourselves to fully-insured plans only while measuring the dependent variable because there is no valid measure of "premiums" for self-insured plans. As can be seen from the final 2 columns of Table 5.6, our results are quite robust to this alternate sample definition as well.

5.5 Discussion

In this paper, we study the effect of local market concentration on premiums charged by private health insurance companies. Using a novel dataset containing information on over 10 million insured lives, we employ instrumental variables techniques to estimate the price-concentration relationship. We use the number of carriers in a market as an instrument for market concentration. The estimates from the IV regression imply a 2.5-4% increase (depending on the measure of concentration employed) in premiums for a one standard deviation increase in market concentration.

While this may not seem that substantial, one should bear in mind the fact that insurers often operate at a medical loss ratio of over 95%, i.e. costs for covering medical claims total 95% of premium revenues collected. Once we account for this fact, we see that this is a sizable effect.

Our analysis does have its share of limitations which we are trying to address in ongoing work. As discussed in section 5.4, the instrument that we use, number of carriers in a market, has its share of weaknesses. We are trying to develop an instrument that would clearly be exogenous with respect to changes in premiums. To that end, we are exploring the idea of using mergers among national insurance carriers as an instrument for market concentration.

Second, we have restricted attention to HMO plans in this study. It is conceivable that HMO premiums are also affected by market structure of indemnity or PPO plans. Finally, we need to take into account the costs faced by the insurance carriers from the provider side. Premium increases may arise not only due to increased insurer market power but could also be due to cost shocks. Our current analysis looks at within-market changes, but does not account for differential growth in costs across different markets. To this end, we plan to use the Medicare Cost Report data to construct an index of hospital prices for this period which can then be aggregated to form market-level proxies for costs faced by insurance carriers.

6 References

Argote, L., Beckman, S. and Epple, D. 1990. "The Persistence and Transfer of Learning in Industrial Settings", *Management Science,* 36(2), 140-154

Argote, L., Brooks, D., and Regans, R. 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

Argote, L., Epple, D., Devadas Rao, R., and Murphy, K. 1997. "The Acquisition and Depreciation of Knowledge in a Manufacturing Organization: Turnover and Plant Productivity", *Working Paper*, Carnegie Mellon University

Argote, L., Epple, D., and Devadas Rao, R. 1991. "Organizational Learning Curves: A Method for Investigating Intra-Plant Transfer of Knowledge Acquired Through Learning-by-Doing", *Organization Science*, 2(1), 58-70

Argote, L., Epple, D., and Murphy, K. 1996. "An Empirical Investigation of the Microstructure of Knowledge Acquisition and Transfer Through Learning-by-Doing", *Operations Research,*44(1), 77-86

Becker, G. 1962. "Investment in Human Capital: A Theoretical Analysis", *The Journal of Political Economy*, 70(5), 9-49

Benkard, L. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production", *American Economic Review*, 90(4), 1034-1054

Besanko, D., Doraszelski, U., Kryukov, Y. and Satterthwaite, M. 2005. "Learning-by-Doing, Organizational Forgetting, and Industry Dynamics", *Working Paper*, Northwestern University

Bessen, J. 1998. "Productivity Adjustments and Learning-by-Doing as Human Capital", *Research on Innovation Working Paper*

Bikhchandani, S., Chandra, A., Goldman, D. and Welch, I., 2001, "The Economics of Iatroepidemics and Quakeries: Physician Learning, Information Cascades, and Geographic Variation in Medical Practice", *Mimeo*

Borenstein, S. and Rose, N. 1994. "Competition and Price Dispersion in the U.S. Airline Industry", *Journal of Political Economy*, 102, 653-683

Bresnahan, T. 1989. "Empirical Studies in Industries with Market Power", *Handbook of Industrial Organization, Vol. II*, edited by Richard Schmalensee and Robert Willig

Cotterill, R. 1986. "Market Power in the Retail Food Industry: Evidence from Vermont", *Review of Economics and Statistics*, 68(3), 379-386

Dafny, L. 2007. "Are Health Insurance Markets Competitive? A Test of Direct Price Discrimination", *mimeo*, Northwestern University

Darr, E., Argote, L. and Epple, D. 1995. "The acquisition, transfer and depreciation of knowledge in service organizations: Productivity in franchises", *Management Science,* 41(11), 1750-1762

Davis, P. 2005. "The Effect of Local Competition on Admission Prices in the U.S. Motion Picture Exhibition Market", *Journal of Law and Economics*, 48, 677-708

Delaney, P., Reder, L., Staszewski, J.J. and Ritter F. 1998 "The Strategy-Specific Nature of Improvement: The Power Law Applies by Strategy within Task", *Psychological Science*, 9, 1-7

DiMaggio, Paul J. and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields" *American Sociological Review* 48: 147-160.

Dranove, D. 1988. "Demand Inducement and the Physician/Patient Relationship", *Inquiry*, 26(2), 281-298

Epstein, Andrew, Jonathan Ketcham, and Sean Nicholson, 2005, "The Welfare Effects of Physician Specialization: Do We Want Physicians to Practice Alike?", working paper

Evans, W., Froeb, L. and Werden, G., 1993, "Endogeneity in the Concentration-Price Relationship: Causes, Consequences and Cures", *Journal of Industrial Economics*, 41(4), 431-438

Farley, D. and Ozminkowski, R. 1992. "Volume Outcome Relationships and in hospital mortality: the effect of changes in volume over time" *Medical Care*, 30(1)

Feldman, R. 1994. "The Welfare Economics of a Health Plan Merger", *Journal of Regulatory Economics*, 6, 67-86

Feldman, R. and Wholey, D. 1996. "Effect of mergers on HMO premiums", *Health Care Financing Review*, 17(3), 171-190

Fisher, E. S., J. E. Wennberg, T. A. Stukel, J. S. Skinner, and S. M. Sharp. 1994. "Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven." New England Journal of Medicine 331 (15): 989-95.

Flood AB and Fennell ML. "Through the Lenses of Organizational Sociology: The Role of Organizational Theory and Research in Conceptualizing and Examining Our Health Care System." *Journal of Health and Social Behavior.* 36:154–169, 1995

Fournier, G., Prasad, K. and Burke, M., 2004, "Physician Social Networks and Treatment Variations in Coronary Inpatient Care", *Mimeo,* Florida State University

Fuchs, V. 1978. "The Supply of Surgeons and the Demand for Operations", *Journal of Human Resources*, 13, 35-56

Gaynor, M., Seider, H. and Vogt, W. 2005 "The Volume-Outcome Effect, Scale Economies and Learning-by-Doing", *American Economic Association Papers and Proceedings*, May 2005, 243-247

Gibbons, R. and Waldman, M. 2004. "Task-Specific Human Capital", *American Economic Association Papers and Proceedings*, May 2004, 203-208

Gowrisankaran, G.., Ho, V. and Town, R. 2006 "Causality and the Volume-Outcome Relationship in surgery", *NBER Working Paper*

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

Gruber, J., Kim, J., Mayzlin, D. 1999. "Physician Fees and Procedure Intensity: The Case of Cesarean Delivery", *Journal of Health Economics*, 18, 473-490

Halm E., Lee C. and Chassin, M. 2002. "Is Volume related to outcome in health care? A systematic review and methodologic critique of the literature", *Annals of Internal Medicine,* 137(6)

Hamilton, B. and Hamilton, V. 1997 "Estimating Surgical Volume-Outcome Relationships Applying Survival Models: Accounting for Frailty and Hospital Fixed Effects", *Health Economics,* 6: 383-395

Hannan, E., O'Donnell, J., Kilburn, H, Bernard, H., and Yazici, A. 1989. "Investigation of the relationship between volume and mortality for surgical procedures performed in NY state hospitals", *JAMA,* 262(4), 503-510

Hannan, E., O'Donnell, J., Bernard, H., Lucacik, G. and Shields, E. 1991. "Coronary Artery Bypass Graft surgery: The relationship between in-hospital mortality rate and surgical volume after controlling for clinical risk factors", *Medical Care,* 29(11), 1094-1107

Hannan, E., O'Donnell, J., Bernard, H., Yazici, A, Lindsey, M.. and Shields, E. 1992. " A longitudinal analysis of the relationship between in-hospital mortality in New York state and the volume of Abdominal Aortic Aneurysm surgeries performed", *Health Services Research,* 27(4)

Harrington, D. and Krynski, K. 2002. "The Effect of State Funeral Regulations on Cremation Rates: Testing for Demand Inducement in Funeral Markets", *Journal of Law and Economics*, 45, 199-225

Hatch, N. and Mowery, D. 1998 "Process Innovation and Learning-by-Doing in Semiconductor Manufacturing", *Management Science*, 44, 1461-1477

Hirshfeld, J.W., Ellis S.G., Faxon, D.P. et al. 1998 "Recommendations for the Assessment and Maintenance of Proficiency in Coronary Interventional Procedures", *Journal of the American College of Cardiology*, 31(3), 722-743

Ho, V. 2002. "Learning and the evolution of medical technologies: The diffusion of Coronary Angioplasty" *The Journal of Health Economics,* 21(5), 873-885

Hubbard, T. 2002. "How Do Consumers Motivate Experts? Reputational Incentives in an Auto Repair Market", *Journal of Law and Economics*, 45(2), 437-468

Huckman, R. and Pisano, G. 2006, "The firm-specificity of individual performance: Evidence from cardiac surgery", *Management Science,* 52(4), 473-488

Hughes, R., Hunt, S. and Luft, H. 1987 "Effects of Surgeon volume and hospital volume on quality of care in hospitals" *Medical Care* 25(6), 489-503

Jovanovic, B. 1979 "Firm-Specific Capital and Turnover", *The Journal of Political Economy*, 87, 1246-1260

Killingsworth, M. 1982 "Learning-by-Doing and Investment in Training: A Synthesis of Two Rival Models of the Life Cycle", *The Review of Economic Studies*, 49(2), 263-271.

Luft, H., Bunker, J.P., Enthoven, A. 1979. "Should operations be regionalized? The empirical relation between surgical volume and mortality". *New England Journal of Medicine.* 301(25), 1364-1369

Luft, H, Hunt, S., and Maerki, S. 1987 "The Volume-Outcome Relationship: Practice-makes-Perfect or Selective-Referral Patterns?" *Health Services Research* 22(2), 157-182

Luft, H., Garnick, D., Mark, D., and McPhee, S. 1990. "Hospital Volume, Physician Volume and Patient Outcomes", Ann Arbor Health Administration Press

McCourt, C., Weaver, J., Statham, H., Beake, S., Gamble, J., Creedy, D. 2007. "Elective Cesarean Section and Decision Making: A Critical Review of the Literature", *Birth*, 34(1), 65-79

McGuire, T., Pauly, M. 1991. "Physician Response to Fee Changes with Multiple Payers", *Journal of Health Economics*, 10, 385-410

McPherson et al., "Regional Variation in the Use of Common Surgical Procedures: Within and Between England and Wales, Canada and the USA," *Social Science and Medicine* Vol 15A, pp 273-288.

Mirkowsky, J., Ross, C. and Reynolds, J., 2000, "Links Between Health Status and Social Status" in Chloe Bird, Peter Conrad and Allan Fremont eds., *Handbook of Medical Sociology, Fifth Edition*, Pp. 47-67.

Nembhard, D. and Uzumeri, M. 2000. "An Individual-Based Description of Learning within an Organization", *IEEE Transactions on Engineering Management,* 47(3), 370-378

Newmark, C. 1990. "A New Test of the Price-Concentration Relationship in Grocery Retailing", *Economic Letters*, 33, 369-373

Newmark, C. 2004. "Price Concentration Studies: There You Go Again", *Manuscript prepared for the DOJ/FTC Workshop*

Nicholson, S. and Epstein, A., "The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections", 2005, NBER Working Paper.

Phelps, CE. 2003. "Health Economics", Third Edition, Addison-Wesley Press

Phelps, C. et al., "Doctors have Styles—And They Matter!" University of Rochester Working Paper.

Phelps, C., 2000, "Information Diffusion and Best Practice Adoption" in *Handbook of Health Economics*, edited by Culyer, A. and J. Newhouse, Amsterdam: Elsevier

Phelps, C. 2003. "What's Enough? What's Too Much"? Annals of Internal Medicine, 18: 4: 348-349.

Phelps, C. and Mooney, C., 1993, "Variations Huin Medical Practice Use: Causes and Consequences", in R.J. Arnould, R.F. Rich and W.White, eds., *Competitive Approaches to Health Care Reform.* Washington D.C. : Urban Institute Press

Phelps, C. and Parente, S., 1990, "Priority Setting for Medical Technology and Medical Practice Assessment", *Medical Care*, 28(8), 703-23

Picone, G., Trogdon, J. and Jollis, J. 2005 "Hospital Volume and Quality of Care: practice-makes-perfect or selective referral", *NBER Working Paper*

Pisano, G., Bohmer, M.J., and Edmondson, A. 2001 "Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery", *Management Science*, 47, 752-768

Pritchard, RS, Fisher ES, Teno JM et al, 1998, "Influence of patient preferences and local health system characteristics on the place of death", *Journal of American Geriatric Society*, 46: 1242-1250

Robinson, J. 2004. "Consolidation and the Transformation of Competition in Health Insurance", *Health Affairs*, 23(6), 11-24

Roemer, M. 1961. "Bed Supply and Hospital Utilization: A Natural Experiment", *Hospitals* 35, 36-42

Schmalensee, R. 1989. "Inter-Industry Studies of Structure and Performance", *Handbook of Industrial Organization, Vol. 1*, edited by Richard Schmalensee and Robert Willig

Scott, R.W., Ruef, M., Mendel, P. and Caronna, C. (2000), "Institutional Change and Health Care Organizations: From Professional Dominance to Managed Care", University of Chicago Press, Chicago.

Scott. W. R. 2001. Institutions and Organizations, 2nd edition. San Francisco, CA : Sage.

Shortell, S. 2004. "Increasing Value: A Research Agenda for Addressing the Managerial and Healthcare Challenges Facing Healthcare Delivery in the United States", Medical Care Research and Review, 61: 3: 12S-30S.

Shafer, S.M., Nembhard, D., and Uzumeri M. 2001. "The Effects of Worker Learning, Forgetting and Heterogeneity on Assembly Line Productivity", *Management Science*, 47, 1639-1653

Soest, A.V., Bartels, R., Fiebig, D. 2006. "Consumers and Experts: An Econometric Analysis of the Demand for Water Heaters", *Empirical Economics*, 31(2), 369-391

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

Thompson, P. 2001. "How Much Did the Liberty Shipbuilders Learn? Evidence for an Old Case Study", *Journal of Political Economy*, 109(1), 103-37

Thompson, P. 2006. "How Much Did the Liberty Shipbuilders Forget?", *Working Paper*, Florida International University

Thornton, R. and Thompson, P. 2001. "Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding", *American Economic Review*, 91(5):1350-1368

Town, R. 2001. "The Welfare Impact of HMO Mergers", *Journal of Health Economics*, 20, 967-994

Wennberg JE, Fisher ES, Stukel TA, Sharp SM, 2004. "Use of Medicare Claims Data to Monitor Provider Specific Performance among Patients with Severe Chronic Illness", *Health Affairs,* Web exclusive, 5-18

Wennberg JE, Gittelson A., 1973, "Small area variation in health care delivery", *Science*, 182, 1102-08

Wennberg JE, Gittelson A., 1982, "Variation in medical care among small areas", *Scientific American*, 246:4: 120-134.

Wholey, D., Feldman, R. and Christianson, J.B. 1995. "The Effect of Market Structure on HMO Premiums", *Journal of Health Economics*, 14, 81-105

Wright, T.P. 1936. "Factors Affecting the Cost of Airplanes", *Journal of the Aeronautical Sciences*, 3(4): 122-128