NBER WORKING PAPER SERIES

DIAGNOSIS AND UNNECESSARY PROCEDURE USE: EVIDENCE FROM C-SECTION

Janet Currie
W. Bentley MacLeod

Working Paper 18977
http://www.nber.org/papers/w18977

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2013

We thank Samantha Heep and Dawn Ko˙man for excellent research assistance, and Amitabh Chandra, Jonathan Gruber, Amy Finkelstein, Jonathan Skinner and seminar participants at the Princeton and the University of Michigan for helpful comments. This research was supported by a grant from the Program on U.S. Health Policy of the Center for Health and Wellbeing. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Diagnosis and Unnecessary Procedure Use: Evidence from C-Section
Janet Currie and W. Bentley MacLeod
NBER Working Paper No. 18977
April 2013
JEL No. I11

ABSTRACT

This paper provides a model of diagnostic skill as an element of provider quality that is separate from a doctor's skill in performing procedures. Unlike higher surgical skill, which leads to higher use of surgical procedures across the board, better diagnostic skill results in fewer procedures for the low risk, but more procedures for the high risk. That is, better diagnostic skill improves the matching between patients and procedures leading to better health outcomes.

Taking the model to data on C-sections, the most common surgical procedure performed in the U.S., we show that improving diagnostic skills from the 25th to the 75th percentile of the observed distribution would reduce C-section rates by 11.7% among the low risk, and increase them by 4.6% among the high risk. Since there are many more low risk than high risk women, improving diagnosis would reduce overall C-section rates. Moreover, such an improvement in diagnostic skill would improve health outcomes for both high risk and low risk women, while improvements in surgical skill have the greatest impact on high risk women. These results are consistent with the hypothesis that efforts to improve diagnosis through methods such as checklists, computer assisted diagnosis, and collaborative decision making may improve patient outcomes.

Janet Currie
Princeton University
316 Wallace Hall
Princeton, NJ 08544
and NBER
jcurrie@princeton.edu

W. Bentley MacLeod
Department of Economics
Columbia University
420 West 118th Street, MC 3308
New York, NY 10027
and NBER
bentley.macleod@columbia.edu

1 Introduction

High and rising health care costs are a major source of fiscal stress in the United States where they already account for 18% of GDP.[1] Unnecessary procedure use is one driver of increasing costs (Garber and Skinner (2008)). This problem has even been recognized by physician groups: The Choose Wisely Campaign unveiled in April 2012 includes nine specialty societies representing 374,000 physicians that have developed checklists and patient-friendly guides aimed at eliminating unnecessary tests and procedures.[2] Many possible reasons have been advanced for unnecessary procedure use including patient demand; defensive medicine (that is, fear of lawsuits); the profit motive; spillover effects on physician practice style; and physician specialization in high tech procedures which may be inappropriate for low risk patients (Chandra et al. (2011)).

This paper focuses on poor diagnostic skill as a reason for inappropriate procedure use. Most previous analyses of physician decision making have focused on a single dimension of physician skill, viz. skill in performing procedures. Instead, we develop a model in which physicians have two dimensions of skill: They may be more or less skilled at doing procedures, and they may be more or less skilled at diagnosis. Diagnostic skill is the ability to reliably transform observed symptoms into an assessment of patient condition. We model a diagnosis as a decision problem in which the physician uses the available information to update her prior beliefs regarding a patient's condition.

Although it has been neglected in the health economics literature, diagnostic skill has become an increasingly important issue because of the growing complexity of medical care and the sheer number of different treatment op-

[1]See https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/proj2010.pdf, accessed Dec. 16, 2012.

[2]See http://www.abimfoundation.org/Initiatives/Choosing-Wisely.aspx, accessed Dec. 16, 2012.

tions available. For example, in a world in which there was little that could be done for cancer patients, it did not matter if physicians correctly diagnosed the disease; now it may be a matter of life or death whether a breast cancer is correctly diagnosed as estrogen-sensitive or not. The increased importance of diagnosis is reflected both in growing attention to medical errors as a leading cause of morbidity and mortality (Committee on Identifying and Preventing Medication Errors (2007)), and in growing numbers of malpractice cases that focus on misdiagnosis (Mello and Studdert (2007)[3]). Finally, an additional reason to focus on diagnosis is that it may be possible to improve it through mechanisms such as checklists, computer aided diagnosis, or administrative structures that support collective decision making (Baker et al. (2008); Doi (2007); Gawande (2009)).

We examine the role of diagnosis in the context of Cesarean section delivery. There is a consensus that there are too many C-sections in the U.S., with rates of 35% vs. the 15% rate that is thought to be closer to optimal. Not surprisingly, the marginal C-section is unnecessary (Baicker et al. (2006)). For our purposes, C-section, which is the most common surgical procedure in the U.S., is ideal because given the detailed records collected for each birth, we can identify women with a high or low risk of C-section a priori, and we can also identify a variety of negative health outcomes following delivery.

We show that improvements in diagnosis increase the incidence of C-sections for high risk women, but reduce the incidence of C-sections for low risk women. Since low risk women outnumber high risk ones, improving diagnosis would be associated with an overall reduction in C-section rates. Specifically, we estimate that moving a woman from a provider at the 25th percentile of the distribution of diagnostic skill to a provider at the 75th percentile would reduce the probability of C-section among low risk women by 11.7%, and would increase the probability of C-section among high-risk women by 4.6%. By way of comparison, increasing providers' skill performing

[3]They find that 70% of malpractice cases are due to errors of judgment.

C-sections by a comparable amount would increase C-section rates by about 4.4% among low risk women, but by only .8% among high risk women.

Moreover, since most low risk women better off without C-sections while most high risk women are better off with C-sections, improved diagnosis reduces the risk of bad outcomes for all women. Our estimates suggest that improvements in diagnosis of the magnitude described above would reduce the incidence of poor outcomes by 13.3% among low risk women, and by 7.3% among high risk women. In contrast, improving surgical skill per se is beneficial for all women, but has much larger effects among high risk women. Finally, improving surgical skill increases the return from performing a C-section for both high and low risk patients, and hence does not directly address the issue of inappropriate procedure use.

The rest of our paper is laid out as follows. Section II provides a brief overview of the existing literature on the reasons for unnecessary procedure use. Section III lays out our model. Section IV provides a description of our data and empirical methods. Results are described in Section V and Section VI concludes.

2 Background

One of the most common explanations for unnecessary procedure use is "defensive medicine", the idea that doctors do unnecessary procedures in order to protect themselves against lawsuits. This view persists despite being debunked by many studies. For example, Baicker et al. (2007) argue that there is little connection between malpractice liability costs and physician treatment of Medicare patients, and Dubay et al. (1999) cast doubt on such a relationship for C-section deliveries.

Currie and MacLeod (2008) conduct a theoretical and empirical examination of the effect of tort reform on the use of C-section. They develop a model in which patients can be ranked in terms of appropriateness for C-section, and show that the doctor's optimal threshold for performing C-section varies with the liability risk. They argue that if doctors are doing C-sections in order to protect themselves from legal liability, then tort reforms that reduced liability should reduce C-section. Instead, they show that reducing liability increases the use of C-section. The intuition is simple: If the marginal C-section is unnecessary, then it is likely to do more harm than good. Reducing the liability from harming people by doing unnecessary surgeries therefore increases the number of such unnecessary surgeries.

Currie and MacLeod's result strongly suggests that doctors have other motives besides fear of lawsuits for performing C-sections. The profit motive is an obvious alternative explanation. The fee for performing C-sections exceeds the fee for performing vaginal deliveries. Moreover, C-sections take less time and can be scheduled at a time that is convenient for doctors. Gruber and Owings (1996) and Gruber et al. (1999) show that the incidence of C-section among Medicaid patients increases with the gap between the fee for C-section and vaginal delivery. However, the profit motive does not provide a complete model of doctor behavior. Since doctors always make more money doing C-sections, a simple profit motive would presumably lead to even higher C-section rates than we already observe.

Hence, researchers have also considered other determinants of doctor behavior including the idea of "practice style" which is often proxied by a physician fixed effect in a model of procedure use. The origins of distinct practice styles remains a mystery: Esptein and Nicholson (2009) use data from Florida and find little evidence of convergence in practice styles over time within hospitals. They further find little effect of the physician's residency program. Dranove et al. (2011) use the same data from Florida to examine the evolution of physician practice styles and find strikingly little evidence of changes over time. They conclude that physicians in the same hospital tend to have similar practice styles because of matching, not because they learn from each other.

Chandra and Staiger (2007) develop a model in which providers specialize in either a high intensity or a low intensity procedure. The specific example they consider is medical management (drugs) vs. surgery for heart attack patients. A key element of their model is that specialization makes doctors better at what they do, but also has an opportunity cost: High intensity providers are better at surgery, but worse at medical management, whereas low intensity providers are better at medical management but worse at surgery. One of the main implications of the model is that patients who are good candidates for surgery will benefit from going to high intensity providers, while patients who are bad candidates for surgery will benefit from going to low intensity providers.

In the Chandra-Staiger world, doctors tend to do what they are good at. We show below that considering diagnostic skill as well as procedural skill yields additional implications. For example, in a world with specialization in high intensity and low intensity procedures, improving the diagnostic skills of a high intensity provider can paradoxically lead to worse outcomes for low risk patients because doctors will do less of the high intensity procedures that they are good at, and more of the low intensity procedures that they are bad at, on these patients. We will show empirically that high risk patients do benefit from going to a provider with excellent procedural skills as Chandra-Staiger predict. However, in contrast to their model, low risk patients do not suffer from going to such a physician. Rather, the low risk patients suffer if they go to a physician with poor diagnostic skills.

Few researchers in economics have considered diagnosis and procedural skill as distinct aspects of medical practice, or attempted to model diagnosis. In a rare exception, Afendulis and Kessler (2007) show that doctors who provide both diagnosis and specialized services are more likely to recommend their own services, which yields overuse, but also some productive efficiencies. We explore the relationship between diagnosis, procedural skill and outcomes more formally below.

3 A Model of Diagnostic and Surgical Skill

3.1 Understanding Physician Decision Making

In this section we discuss how to use the standard Roy model of physician decision making to understand physician diagnosis.[4] In our data we observe patient characteristics, the procedure chosen by the physician, and various measures of medical outcomes. The goal is to understand how variations in physician skill affect procedure use and medical outcomes. In particular we explore how variations in a physician j's ability to process information is likely to impact procedure choice and performance.

3.2 Information

Diagnosis means the ability to reliably transform observed symptoms into an assessment of patient condition. The treatment depends on the diagnosis, but also on the costs of treatment, the doctor's skill in performing procedures, and on patient preferences. To the extent that diagnosis affects the course of treatment it can lead to better or worse outcomes. In our data we cannot observe all the information that is available to the physician, but we do have a very rich set of observed conditions, X_i, for patient i. We also observe the procedure that doctor j chooses for patient i, denoted by $T_{ij} \in \{N, C\}$, where N and C represent the non-intensive and intensive procedures, corresponding to natural delivery and a C-section in our data. The model we discuss can be applied to any situation where the physician faces a dichotomous choice.

In order to compare diagnosis across physicians, we begin by creating a measure of patient appropriateness for procedure $T_{ij} = C$. We estimate a discrete choice model:

$$\rho_i^r = F\left(\beta^r X_i\right), \tag{1}$$

where F is either a logistic or normal distribution, and $\rho_i^r \in [0,1]$ is the

[4] This is the model used in Chandra and Staiger (2007) and Currie and MacLeod (2008).

predicted probability of procedure C. Let $s_i^r = \beta^r X_i \in \Re$ be the corresponding index that varies over the real line. We show that ρ_i^r indexes a physician-independent measure of the appropriateness of the patient receiving treatment $T = C$ that does a good job assessing an individual patient's need. This measure is an average over the whole market, hence any individual physician's contribution to ρ_i^r is very small.

Our goal is to understand both treatment choice, and the impact upon patient welfare. We approach this problem by supposing that there is an underlying state of the patient, $s_i \in \Re$, with the interpretation that this is the *net medical benefit* of doing procedure C, and that C should be carried out whenever the benefit is positive or $s_i \geq 0$. Thus, we can interpret $s_i^r = \beta^r X_i$ as the market's best estimate of the patient's condition, and we will assume that it forms a proxy for the net benefit of procedure C.

3.3 Physician Behavior

We follow the literature, and suppose that the physician chooses the best action possible given her information, costs, and patient preferences. When deciding upon a procedure the physician evaluates the condition of the patient to produce two latent variable s_{ij}^N and s_{ij}^C which are the predicted outcomes if procedure N or C are performed. The physician's costs are assumed to be affected by both their skill at performing the two procedures, and by the fees they can charge. Hence, the payoff of the physician is:

$$U_{ij}(T) = s_{ij}^T + skill\left(T, S_j^T\right) + m_j^T\left(P^T\right), \tag{2}$$

where $skill\left(T, S^T\right)$ is the skill at doing procedure T and $m_j\left(P^T\right)$ is the expected pecuniary consequence of this choice as a function of the price paid, P^T for procedure T. It is assumed that an increase in physician skill increases the benefit of the procedure:

$$\partial skill(T, S_j^T)/\partial S_j^T \quad > \quad 0.$$

We have rough proxies for price, and suppose that the gain physicians respond to is the difference in price for the two procedures:

$$m_j\left(\Delta P\right) = m_j^C\left(P^C\right) - m_j^N\left(P^N\right),$$

where $\Delta P = P^C - P^N$ is the price difference between a procedure C and natural delivery. The function $m_j\left(\Delta P\right)$ is assumed to be strictly increasing in ΔP.

The medical benefit of a procedure is given by

$$MB_{ij}^T = s_{ij}^T + skill\left(T, S_j^T\right).$$

And the net medical benefit is given by:

$$\begin{aligned} NMB\left(s_{ij}, S_j^C, S_j^N\right) &= MB_{ij}^C - MB_{ij}^N && (3) \\ &= s_{ij} + \Delta skill\left(S_j^C, S_j^N\right) && (4) \end{aligned}$$

where $s_{ij} = s_{ij}^C - s_{ij}^N$, $\Delta skill\left(S_j^C, S_j^N\right) = skill\left(C, S_j^T\right) - skill\left(N, S_j^T\right)$.

Hence the physician carries out a C-section if and only if the net medical benefit plus the pecuniary benefit is positive:

$$NMB\left(s_{ij}, S_j^C, S_j^N\right) + m_j\left(\Delta P\right) \geq 0, \tag{5}$$

The probability that patient i has a procedure C with physician j is given by:

$$\rho_{ij} = F\left(\left(NMB\left(s_{ij}, S_j^C, S_j^N\right) + m\left(\Delta P\right)\right)\right). \tag{6}$$

This expression, combined with our assumptions regarding skill and price implies the following well known result for the Roy model of physician behavior:

Proposition 1. *The rate of procedure C use increases with physician skill in doing procedure C (S_j^C) and with the difference in price between procedure C and procedure N. Procedure C rates fall with an increase in skill in procedure N (S_j^N).*

This result is true both on average for the whole population and conditional upon the patient's risk for having procedure C. Next we consider the issue of diagnostic skill.

3.4 Understanding Diagnosis

As discussed above, we construct a measure of patient condition given by $s_i^r = \beta^r X_i$. This index is constructed using data for the whole state, and hence for any given individual the index is independent of physician characteristics.

We already know that different physicians often make different decisions with the same data regarding a women's condition (Esptein and Nicholson (2009)). This may be in part because they differ in the way that they process information.

We formally capture this effect by supposing that s_i^r , given by equation (1), is an unbiased signal of the net benefit with variance σ_r^2. The physician is assumed to observe:

$$s_{ij} = s_i^r + \epsilon_{ij},$$

where the variance of ϵ_{ij} is σ_{ij}^2, and $D_j = \frac{1}{\sigma_{ij}}$ is the precision of this signal, and hence a measure of *diagnostic skill*.[5] In terms of equation (1), we are assuming that everyone observes the same $X_i$but that doctors use their personal experiences to form β^r. Since we use data for the entire state over 10 years, we are assuming that we have a superior estimate of β^r. The case in

[5]Normally the precision is the reciprocal of the variance σ_{ij}^2, but the reciprocal of the standard deviation σ_{ij}^2 provides a more convenient measure of diagnostic skill.

which doctors observe additional data that we do not observe is considered in section 3.5.2 below.

This structure follows from a rational choice framework in which doctor experiences lead them to have prior beliefs regarding the benefit of procedure C for the average patient. Let s_j^0 be the mean and σ_{ij}^{02}be the variance of these beliefs. Let$D_j^0 = \frac{1}{\sigma_{ij}^0}$ be a measure of how strongly a physician holds his or her pre-existing beliefs.

From DeGroot (1972) we have the familiar learning rule:

$$\begin{aligned} E\left\{s_i|s_{ij}, s_j^0\right\} &= \pi s_{ji} + (1-\pi) s_j^0 \\ &= \pi s_i^r + (1-\pi) s_j^0 + \pi \epsilon_{ij}, \end{aligned} \tag{7}$$

where $\pi = \frac{D_j^2}{D_j^2 + (D_j^0)^2}$.

The point here is that the sensitivity of the updated beliefs to the observed signal is a function of how much information is extracted from X_i.

This expression allows us to put a bit more structure on the decision function 6. If the physician can observe s_i^r directly, then D_j is zero and diagnosis is not an issue. Procedure C is choosen if and only if:

$$s_i^r \geq -\Delta skill\left(S_j^C, S_j^N\right) - m\left(\Delta P\right).$$

This rule is illustrated in Figure 1a where $\bar{\rho}_j = F\left(-\Delta skill\left(S_j^C, S_j^N\right) - m\left(\Delta P\right)\right)$. That is, the doctor determines a threshold patient condition. Only patients with risk above the threshold level receive a C-section. The threshold shifts down (indicating that more C-sections will be performed) whenever C-sections become more lucrative or the doctor's skill in performing C-section increases relative to his or her skill performing natural deliveries. Thus increases in prices for C-section and surgical skill have their greatest impact on the marginal patients.

3.5 Effect of Diagnosis on Decisions and Outcomes

Let us now consider the situation when the doctor doesn't perfectly observe patient appropriateness. Let I_{ij} denote all the information that a physician has when she decides what procedure to perform on patient i. Now, instead of observing the patient's condition, the physician has an expectation about that patient's condition given the information set. A physician will choose to perform C if and only if:

$$E\left\{s_i|I_{ij}\right\} + \Delta skill_j + m_j\left(\Delta P\right) \quad \geq \quad 0. \tag{8}$$

Here we are assuming the physician understands her skill and the pecuniary gains from performing procedure C. Thus her information is only used to make an assessment of patient condition, which is given by $E\left\{s_i|I_{ij}\right\}$. This expected value is solved using Bayes' rule (7) to get:

$$\pi s_i^r + (1-\pi)\, s_j^0 + \Delta skill_j + m_j\left(\Delta P\right) \quad \geq \quad \pi\epsilon_{ij}.$$

If we divide by the weight π we get the expression:

$$s_i^r - a_j \quad \geq \quad \epsilon_{ij} \tag{9}$$

Where

$$a_j = -\frac{(1-\pi)\, s_j^0 + \Delta skill_j + m_j\left(\Delta P\right)}{\pi} \tag{10}$$

is a physician specific constant. Let the probability that a patient i with observed condition s_i^r who is treated by physician j receives procedure C be denoted by ρ_{ij}. Since the the error term ϵ_{ij} is normally distributed with mean zero, and variance $1/D_j^2$ from (9) we have:

$$\begin{aligned}\rho_{ij} &= Prob[s_i^r - a_j \geq \epsilon_{ij}], \\ &= F\left(D_j\left(s_i^r - a_j\right)\right). \end{aligned} \tag{11}$$

For notational simplicity we write ρ_{ij} rather than showing explicitly that it depends upon patient and physician characteristics. In subsequent expressions it is understood that ρ_{ij} can vary with any patient i or physician j characteristic. In the standard Roy model, as used for example by Chandra and Staiger (2007) and Esptein and Nicholson (2009), they assume that only the constant term a_j varies across regions, but that the slope (diagnostic skill), D_j, is constant. One contribution of our work is to explore the implications of allowing D_j to vary between doctors.

Previous work has shown that an increase in surgical skill leads to higher procedure rates, which is also the case in our model:

$$\frac{1}{f\left(D_j\left(s_i^r - a_j\right)\right)} \frac{\partial \rho_{ij}}{\partial \Delta skill_j} = \frac{D_j}{\pi} > 0. \tag{12}$$

However, in our model the size of this derivative varies with diagnostic skill (and also with practice style which comes in via π which depends on D_j^0). Since $\frac{D_j}{\pi}$ increases with diagnostic skill, utilization increases with skill at a faster rate when there is greater diagnostic skill.

We can also derive the effect of diagnosis upon procedure use holding skill, prices and practice style fixed. Taking the derivative of 11 with respect to diagnostic skill we get:

$$\frac{1}{f\left(D_j\left(s_i^r - a_j\right)\right)} \frac{\partial \rho_{ij}}{\partial D_j} = s_i^r - b_j, \tag{13}$$

where b_j is the intercept term plus it's elasticity with respect to diagnostic skill:

$$b_j = a_j + D_j \frac{\partial a_j}{\partial D_j}. \tag{14}$$

The elasticity of the constant term a_j with respect to diagnosis is:

$$D_j \frac{\partial a_j}{\partial D_j} = \left\{ \left((1-\pi)^2 + \pi \right) s_j^0 + \Delta skill_j + m_j (\Delta P) \right\} \frac{2(1-\pi)}{\pi}. \tag{15}$$

This derivative is ambiguous in sign. In general $1 > \pi > 0$ which means that the derivative is positive if and only if:

$$s_j^0 \geq - \left(\frac{\Delta skill_j + m_j (\Delta P)}{\left((1-\pi)^2 + \pi \right)} \right). \tag{16}$$

However, given that the value of b_j does not vary with the condition of the patient and s_i^r can take any real valued expression 13 implies:

Proposition 2. *The probability that the physician uses procedure C increases with diagnostic skill if and only if $s_i^r > b_j$.*

This expression implies that high risk patients will experience an increase in the use of C-section when the physician has better diagnostic skills, and low risk patients will experience decreases in the use of C-section with increases in diagnostic skill.

Propositions 1 and 2 are illustrated in Figures 1b and 1c. In these figures, the probability of C-section rises with patient appropriateness, but it rises more smoothly than in Figure 1a reflecting uncertainty about the actual state of the patient. In Figure 1b an increase in surgical skill or price increases procedure use everywhere (Proposition 1). In contrast, Figure 1c shows that a change in diagnostic skill causes the relationship between C-section and appropriateness to twist and to approach the decision rule given in Figure 1a. These results illustrate that it is possible to disentagle diagnostic skill from surgical skill. An increase in surgical skill should result in an increase in C-sections for all patient types; in contrast, an increase in diagnostic skill increases C-sections for the high risk and reduces them for the low risk.

3.5.1 Outcomes

For high risk patients, the effect of physician characteristics upon the C-section rate is small since most of these patients both need and receive a C-section. Thus, we can use variations in medical outcomes among these patients as a proxy for S_j^C. Similarly we can use outcomes for low risk cases as a proxy for S_j^N (since most low risk patients have natural deliveries). The use of these proxy measures allows us to examine the effect of procedural skill on the physician's propensity to perform C-sections.

Next, let us consider the effect of diagnostic skill, as given by D_j, the precision of the measure of the patient's condition. Our analysis is done in terms of the net medical benefit of C-section relative to natural delivery, which we assume is given by:

$$s_i^r + \Delta skill_j.$$

The physician observes a signal s_{ij} and decides on the procedure following rule 9. We can write the net medical benefit as function of observed medical appropriateness for procedure C as:

$$\begin{aligned} NMB_j\left(s_i^r\right) &= \rho_{ij}\left(s_i^r + \Delta skill_j\right) \\ &- \left(1-\rho_{ij}\right)\left(s_i^r + \Delta skill_j\right) \\ &= \left(2\rho_{ij}-1\right)\left(s_i^r + \Delta skill_j\right) \end{aligned} \tag{17}$$

Hence, the effect of diagnostic skill upon net medical benefit is given by:

$$\frac{\partial NMP_j}{\partial D_j} = 2\frac{\partial \rho_{ij}}{\partial D_j}\left(s_i^r + \Delta skill_j\right)$$

Recall that s_i^r takes values over the whole real line. When s_i^r is sufficiently large then $\frac{\partial\rho_{ij}\left(s_i^r\right)}{\partial D_j} > 0$, and the term $(s_i^r + \Delta skill_j)$ is positive; hence diagnositic skill has a positive effect upon outcomes. Similarly, when s_i^r is sufficiently small, $\frac{\partial\rho_{ij}\left(s_i^r\right)}{\partial D_j} < 0$, and the term $(s_i^r + \Delta skill_j)$ is negative, and hence the total effect is still positive. These results suggest that when patients are either high risk or low risk, improvements in diagnosis will make patients better off. For patients of medium risk, diagnosis interacts with other factors to affect patient outcomes. For example, if a doctor is much better at doing C-sections than natural deliveries, and too many C-sections are being done, then improvements in diagnosis could conceivably make the patient worse off.

The effect of surgical skill on outcomes is given by:

$$\frac{\partial NMP_j}{\partial \Delta skill_j} = 2\frac{\partial \rho_{ij}}{\partial \Delta skill_j}\left(s_i^r + \Delta skill_j\right) + \left(2\rho_{ij} - 1\right).$$

Better surgical skill (relative to natural delivery) always increases the number of C-sections. For high risk patients, $s_i^r > max\left\{a_j, -\Delta skill_j\right\}$, so both $(s_i^r + \Delta skill_j)$ and $(2\rho_{ij} - 1)$ and positive. Hence, the effect of skill is positive. We have a negative sign when $s_i^r < min\left\{a_j, -\Delta skill_j\right\}$, and hence skill has a negative effect on net benefits for low risk patients. Again, there is some indeterminacy about the sign for those at medium risk for whom it is not clear which term predominates.

These effects are illustrated in Figure 2. The figure shows that the marginal benefit from increased diagnostic skill is U-shaped in patient appropriateness for C-section, and that it is positive for patients at both low risk and high risk of C-section. In the middle, the sign of the effect is indeterminant (and it is relatively small). That is, for cases that are marginal medically, it will not do too much harm to make the "wrong" decision. In contrast, the benefit from increased surgical skill (relative to skill at natural deliveries) is increasing in patient appropriateness, and is highest for high risk cases.

Proposition 3. *The effect of diagnostic skill, surgical skill and price on medical outcomes is summarized in the following table:*

	Appropriateness for Procedure C		
	Low	Middle	High
Diagnostic Skill	+	?	+
Surgical Skill	-	?	+

In the standard Roy model increases in surgical skill can lead to some mismatch between the patient and proceedure, an effect highlighted by Chandra and Staiger. Here we show that this effect can be offset by an increase in diagnostic skill which increases match quality for most patients. The effect is ambiguous for the marginal cases, but these are also the cases for which both procedures have similar benefits, and hence errors in diagnosis would have a small effect. As a consequence we would expect that on average an increase in diagnostic skill would improve outcomes.

An explicit policy instrument is procedure price. In this case the effect is quite straightforword and given by:

$$\frac{\partial NMP_j}{\partial \Delta P_j} = 2\frac{\partial \rho_{ij}}{\partial \Delta P_j}\left(s_i^r + \Delta skill_j\right).$$

An increase in price always increase the rate of procedure C, hence it improves outcomes if and only if $s_i^r + \Delta skill_j > 0$. In other words, for high risk patients an increase in the price of procedure increases the use and hence makes individuals better off. The converse is true for low risk patients.

3.5.2 Alternative Information Structures

We have assumed that not all doctors interpret patient conditions X_i in the same way. That is, different doctors have different values of β. An alternative assumption is that all doctors interpret X_i in the same way but some doctors observe additional information and variations in decisions are due to the additional information that is collected. In this alternative scenario, a Bayesian decision maker would put less weight upon X_i as she acquired additional information. This in turn would imply that sensitivity to s_i would *decrease* with improvements in a physician's diagnostic skills. Recall that in our model a sensitivity to s_i is captured by the slope term, D_j. Hence, this alternative scenario implies that decreases in D_j would improve outcomes.

As we show below, we find exactly the opposite result, suggesting that our characterization of doctor decision making is more realistic and that doctors do not all use the information contained in our measures of patient condition, X_i, with equal efficiency. Another way to think about this issue is to reflect on the fact that the β's we estimate reflect the combined experience of all physician's in New Jersey over a ten year period, whereas any individual doctor has much less experience and hence may be less able to infer the correct β's.

4 Data and Methods

The data for this project come from approximately 1.1 million Electronic Birth Certificates, (EBC) spanning 1997 to 2006, from the state of New Jersey. These records have several important features from our point of view. First, they include detailed information about the medical condition of the mother which enables us to predict, with a fair degree of accuracy, which mothers are likely to need C-sections. In particular, we know the mother's age, whether it is a multiple birth, whether the mother had a previous C-section, whether the baby is breech, whether there is a medical emergency such as placenta previa or eclampsia which calls for C-section delivery, and whether the mother had a variety of other risk factors for the pregnancy such as hypertension or diabetes. Second, we know the method of delivery, including whether a C-section was planned or not.

Third, the birth records include unusually detailed information about birth outcomes. Birth records usually record information about complications of labor and delivery. Infant deaths are of particular interest, but are thankfully rare. When we look at deaths, we focus on neonatal deaths (deaths in the first 30 days) as these are more likely than later deaths to be caused by events at the delivery. In addition to these measures, the New Jersey data also includes information about late maternal complications such as fever and hemorrhage that occur after the delivery. In most of our analyses we will combine these measures and look at the probability that there was "any bad outcome." Our comprehensive measure of bad outcomes includes late maternal complications, neonatal death, selected complications of labor and delivery (excessive bleeding, fever, seizures) and selected abnormal conditions of the infant (brachoplexis, fracture, meconium, birth injury, neurological damage in full term infant). We did not include neurological damage in preterm infants as this might be a result of prematurity itself rather than events at the time of the birth.

Fourth, the data has information about the latitude and longitude of each woman's residence, as well as codes for doctors and hospitals. We found, as a practical matter, that very few doctors practiced in more than one hospital in a single year, hence the choice of doctor also defines the choice of hospital. In total, our data include 71 hospitals and 5,273 birth attendants. Of the birth attendants, 603 were midwives. In our analysis, we focus on doctors since only doctors can perform C-sections.

Finally, the data includes demographic information such as race, education, marital status, and whether the birth was covered by Medicaid which have been shown to be related both to the probability of C-section and to birth outcomes.

We use these data to construct analogs of the key concepts in our model. We define ρ_i^r , the mother's risk of C-section, by estimating a logit model of the probability of C-section given all of the purely medical risks recorded in the birth data, as in equation (1). The model we estimated is shown in Table 1. Table 1 shows that the model predicts well, with a pseudo R-squared of almost .32.

Figure 3 provides another way of gauging the accuracy of the model's predictions. It shows that those who did not have a C-section generally had values of ρ_i^r less than .5, while those with C-sections generally had values of ρ_i^r greater than .5. More particularly, the figure shows that those who had values of ρ_i^r less than .2 were very unlikely to have C-sections, while those with ρ_i^r greater than .8 were highly likely to have C-sections. In what follows, we will designate these two groups as the "low risk" and the "high risk" respectively, and consider those with values of ρ_i^r between .2 and .8 as "medium risk". Of the women deemed high risk, 89% received a C-section, while among the women deemed low risk only 11% received a C-section.

For a given level of medical risk, the probability of C-section increased over our sample period at all but the highest risk levels as shown in Appendix Figure 1. In fact, at the start of our sample period, New Jersey, with a rate of 24%, had a lower C-section rate than several other states, including Arkansas, Louisiana, and Mississippi, while by the end of our sample period, New Jersey had pulled ahead to have the highest C-section rate of any state, at almost 40%. Figure 4 shows that this increase was not due to a change in the underlying distribution of medical risks. The figure shows only a slight increase in the number of high risk cases, which is attributable to an increase in the number of older mothers, mothers with multiple births, and increasing numbers of women with previous C-sections (itself driven by the increasing C-section rate).

Since ρ_i^r is a device for ranking women according to their medical risk, the level is less important than the ordering. We have experimented with alternative models and found that the correlation between the ranking produced by our model, and the ranking produced by several alternative models is above .95. These alternative models included a model that included fewer risk factors, a model that used births from 1997-1999 only, and a model that used only doctors who were below the 25th percentile in terms of the fraction of births with negative outcomes in their practices.

It remains to define measures of diagnostic skill, procedural skill, and prices. In the model, diagnostic skill is captured by the variable D_j. An empirical analog can be obtained for each doctor by using the estimated β's from (1) to create the index of maternal condition s_i^r (this is simply $\beta^r X_i$) and then estimating a regression model for each doctor's propensity to perform C-sections as a function of s_i^r. The coefficient on s_i, denoted by $DiagSkill_j$, is an indicator of how sensitive the doctor is to this index of observable indicators of patient risk and thus captures diagnostic skill.

We measure procedural skill by first calculating the rate of bad outcomes among low risk births, and the rate of bad outcomes among high risk births for each doctor, and then taking the difference between them. This measure is a good proxy for skill because, as noted above, most high risk women get C-sections and most low risk women do not. At the same time, because the high risk and low risk groups are defined only in terms of underlying medical risk factors, the measure is not contaminated by the endogeneity of the actual choice of C-section. This measure is less than zero since bad outcomes are less likely for the low risk than the high risk, but it becomes larger as the rate of bad outcomes falls among the high risk (i.e. with greater surgical skill).

For prices, we use data from the Health Care Utilization Project (HCUP). HCUP includes hospital charges for every discharge. For each doctor, we take the mean price of all C-section deliveries that did not involve any other procedures, less the mean price of normal deliveries without other procedures. This differential varied from \$2,250 to \$8,490 real 2006 dollars, with a median of \$4,756.[6]

Having constructed these measures, we estimate models of the following

[6]It is important to note that physician charges are generally separate from hospital charges. In using this measure, we are implicitly assuming that physicians who practice in expensive hospitals charge more.

form:

$$Outcome_{ijt} = f(DiagSkill_j, \Delta Skill_j, \Delta P_{jt}, X_{it}, month, year), \tag{18}$$

where $Outcome_{ijt} \in \{0, 1\}$, where 0 is a Natural delivery and 1 is a C-Section, i indexes the patient, j indexes the doctor, and t indexes the year. We include month and year effects in order to control for seasonal differences in outcomes and for longer term trends affecting all births in the state (e.g. due to other improvements in medical care). The standard errors are clustered at the level of the physician in order to allow for unobserved correlations across a physician's cases.

Sample means are shown in Table 2. The estimation sample is slightly smaller than in Table 1 because we exclude births that were not attended by a doctor, as well as those for whom we cannot calculate our measure of diagnostic skill (because there are too few births per provider).[7] The first panel shows how the outcome variables vary across the low, medium, and high risk groups. As expected, high risk women have more C-sections, a higher risk of a bad outcome than low risk women, and higher neonatal death rates.

The second panel explores the characteristics of doctors and provides some initial evidence with regard to an important question: The extent to which high risk patients see doctors with particular characteristics. The average doctor in our sample performed about a 1,000 deliveries and this is quite similar across risk groups. Similarly, there appears to be little difference across groups in the measures of diagnostic skill, procedural skill differentials, or price differentials. Nor, perhaps surprisingly, is there much of a difference in the fraction of the doctor's patients who have bad outcomes. That is, although high risk women are more likely to have bad outcomes, there is no evidence that they are likely to see doctors who have either high or low

[7] We also exclude a very small number of doctors who did not have at least one high risk patient and at least one low risk patient.

fractions of patients with bad outcomes in their practices. The one measure that shows some variation across risk groups is the fraction of high risk patients in the practice. It appears that high risk patients are more likely to go to doctors who specialize in high risk patients. Hence, we will estimate models controlling for this variable below.

Finally, the third panel of the table provides an overview of selected maternal and child characteristics including race and ethnicity, maternal education, marital status, and whether the birth is covered by Medicaid. The table suggests that low risk women are disproportionately young and minority women giving birth for the first time whereas women at high risk for C-section tend to be older, married women with second or higher order births.

The main empirical difficulty involved in estimating (18) is that women choose their doctors. If women with high risk pregnancies choose better doctors, then the estimated effect of doctor skill on birth outcomes will be biased towards zero. If, on the other hand, high risk women go to less skilled doctors, then estimates of (18) will overstate the beneficial effect of skill on birth outcomes. We address this selection problem in several ways.

First, Table 2 suggested that high risk women were not more likely to choose doctors who were highly skilled. Table 3 expands on this investigation by presenting correlations between the probability of C-section (ρ_i^r) and doctor characteristics. Correlations are presented for all women, and within risk category. Table 3 confirms that high risk women do not seem to be choosing doctors on the basis of our measures of diagnostic skill or procedural skill. There is however, again some evidence that high risk women choose doctors who specialize in high risk cases. Table 3 also shows that there is a positive correlation between diagnostic skill and surgical skill, though it is a modest .215 to .235. And there is a negative correlation between the rate of bad outcomes in a practice and our two skill measures, which is reassuring.

The other ways that we address the selection issue are as follows: (1) We estimate models with and without controls for the share of high risk patients in the practice since this is the one attribute of doctors that appears to be related to patient selection; (2) we estimate our models excluding planned C-sections (C-sections where there was no trial of labor). The logic behind this test is that women who know that they will have a C-section may have a stronger incentive to select a good surgeon; and (3) we estimate models defining provider characteristics at the market level rather than at the doctor level, which will help if markets are less selected than individual doctors within those markets.

Following Kessler and McClellan (1996) we define a hospital market using the hospitals actually selected by women in a particular zip code in a particular year. Specifically, we include all hospitals within ten miles of the woman's actual residential location, plus any hospital used by more than three women from her zip code of residence.[8] Thus, there is a distinct market, or set of hospital choices, facing each woman at the time of each birth.

Figure 5 shows the distribution of hospitals and illustrates this way of defining markets. The figure shows that most women choose hospitals that are close by, but that some women bypass nearby hospitals in favor of hospitals further away. In some cases, these are regional perinatal centers which are better equipped to deal with high risk cases. For example, women from Princeton New Jersey could give birth in the hospital in town, but many travel as far away as Morristown (two counties to the north) to deliver in other hospitals.[9]

Finally, a fourth way to deal with selection which is of some independent interest is to estimate a model of patient demand for doctors using the Alter-

[8]In the crowded northern New Jersey hospital market, we included only hospitals within five miles of the zip code centroid.

[9]The figure also illustrates that the common practice of drawing a circle around a location in order to define a market is likely to be seriously misleading: A circle wide enough to include all the hospitals actually chosen would include hospitals that were never chosen, and a circle wide enough to include most hospitals could miss specialty hospitals that were further away and yet within the choice set.

native Specific Conditional Logit model (McFadden, 1974). The question we address is whether patients are selecting providers on the basis of the skill measures we have constructed, or some other variable such as distance to hospital.

5 Results

Table 4 shows estimates of equation (18), where the dependent variable is whether there was a C-section. Table 4 indicates that diagnostic skill and procedural skill have quite distinct effects. In markets where providers are relatively good at C-section, all women are more likely to have C-sections. However, better diagnosis significantly reduces the probability of C-section for low risk women, increases the probability for medium risk, and has an especially large positive effect for high risk women. A larger price gap between C-sections and natural deliveries increases C-sections for low risk women, but has the largest effect for women at medium risk as the model predicts. The intuition is that price is more likely to be determinative when the medical case is close to the margin.

One useful way to think about the magnitudes of these effects is to consider moving a woman from a doctor at the 25th percentile of the relevant measure to a doctor at the 75th percentile and then compute percentage changes using the mean C-section rates from Table 2. For the index of diagnostic skill, this movement (of .215 units) would reduce C-section rates by 11.7% among the low risk, but would increase them by 3.1% and 4.6% among medium and high risk women respectively. For the index of procedural skill, this movement (of .05 units) would increase the probability of C-section by 4.4%, 1.5%, and .8% for low, medium, and high risk women, respectively. Since there are many more low risk women (475,270) than high risk women (121,631), these figures imply a large overall decrease in C-section rates with better diagnosis. Specifically, they imply a net decrease of 50,012 women receiving C-sections, which is about 5% of the births over our sample period. Finally, the estimates imply that a one standard deviation increase (about $2,600) in the gap between prices for C-section and normal delivery would increase C-section rates by 8.1% among the low risk and 3.3% among the medium risk but would have no impact on the high risk, where medical necessity is a much more important factor than price.

Table 4 also shows the coefficients on the measures of personal characteristics that are included in our models. Most of these characteristics have statistically significant effects on the probability of C-section. As a group, they tend to belie the idea that C-sections are demanded by white, college-educated women. Instead, it appears that African-American and Hispanic women are more likely to have C-sections, as are less educated women. We also see that married women are less to have C-sections while those on Medicaid are more likely.

Table 5 examines birth outcomes. Recall that while the model implies that C-sections decrease for the low risk and increase for the high risk, better diagnosis is predicted to improve outcomes for everyone. Table 5 shows that this is in fact the case. An improvement in diagnosis that moved the doctor from the 25th to the 75th percentile of the distribution would reduce the incidence of any bad outcome by 13.3%, 12.1%, and 7.3% among the low, medium, and high risk, respectively. The incidence of neonatal death also declines by 31.4%, 49.2%, and 45.1% in the same groups (though since neonatal deaths are a rare outcome, these percentage changes should be taken with a grain of salt). Improvements in surgical skill relative to skill doing normal deliveries is also estimated to improve outcomes: Changing from a provider at the 25th percentile of the procedural skill distribution to one at the 75th percentile would be associated with reductions of 7.4%, 18.5%, and 65.4% in the probability of a bad outcome, suggesting especially large effects of surgical skill for the difficult cases. The corresponding estimates for the effects of improvements in surgical skill on neonatal death are 7.3%, 21.9%, and 58.7%. Finally, an increase in the price gap of $2,600 is estimated to increase the risk of neonatal death among the low and medium risk (by 12.3% and 11.9% respectively) but to have no effect on the risk of death among the high risk. This later result is consistent with the evidence that the choice of procedure is not affected by price in the high risk cases.

5.1 Accounting for Selection

To this point, we have ignored the possible impact of doctor selection on our estimates. As discussed above, if women with difficult cases are more likely to choose skilled doctors, then we will tend to under-estimate the effects of skill on outcomes. High risk women being matched with the least skilled doctors, the opposite type of selection, is a more serious potential problem as it has the potential to generate spurious effects of skill. While there is no perfect answer to this selection problem, in this section we explore several alternative estimation strategies.

Table 3 suggested that the main observable difference between doctors treating low risk and high risk patients is that the later are more likely to specialize in high risk patients. Accordingly, in Table 6, we add this observable characteristic of doctors to the model. Controlling for the share of high risk patients in the practice has very little effect on the estimated coefficients on the other doctor characteristics. Specialization itself is associated with a higher probability of C-section, especially among the medium risk group, and with a higher probability of bad outcomes. This later result could reflect the selection we are trying to account for: If high risk women are both more likely to have bad outcomes and more likely to see doctors who specialize in high risk patients, then we would expect this effect. Appendix Table 1 shows that the results are quite similar if we break the share high risk in the practice into quartiles and include those rather than the continuous measures.

Table 7 shows the results of a second experiment in which we exclude planned C-sections from the sample on the grounds that women planning to have a C-section may be more selective in their choice of physician than those who are not. Comparing the first three columns of Table 7 to Table 3 indicates that the effect of diagnosis on the probability of C-section is affected by the exclusion of planned C-sections. Among the low and medium risk, the planned C-sections may be the cases where some diagnostic criterion that we do not observe dictates a C-section. In these two groups we find that the effect of diagnostic skill is reduced by the exclusion of planned C-sections, though the effect remains significantly negative in the low risk group. In the high risk group, the estimated effect of diagnostic skill is much higher when planned C-sections are excluded.

In contrast, the estimated effects of procedural skill on the incidence of C-section are not much affected, and the estimated effect of the price gap is reduced, suggesting that planned C-sections are more sensitive to price than unplanned C-sections. Comparing the remaining columns of Table 7 to Table 5 indicates that excluding planned C-sections has little impact on the estimated effects of diagnostic skill, procedural skill, or price on bad outcomes.

Table 8 shows the results of estimating models where the measures of diagnostic skill, procedural skill, and price are calculated at the market level. As discussed above, a market includes nearby hospitals as well as all of the hospitals in which at least three women from the index woman's zip code delivered in a given year. Since the type of medical services could be correlated with other characteristics of residential location, we include controls for the zip code of residence in these models. Hence, the implicit assumption in these models is that women do not choose their residence on the basis of year-to-year changes in the type of medical services offered in the area. We also cluster the standard errors at the zip code level.

In these market-level models, diagnostic skill is measured using the second proxy discussed in the model section: The difference between the risk adjusted C-section rate for high risk patients and the risk adjusted C-section rate for low risk patients.[10] In order to compute this measure, we take the mean residuals from (1) for high risk patients in the market, and the mean residuals for low risk patients in the market and subtract. This measure has a mean of -0.011 in the whole sample and increases when either the C-section rate for high risk patients increases or when the C-section rate for low risk patients falls.

The measure of the procedural skill differential is defined as it was above (the incidence of poor outcomes for low risk patients in the market minus the incidence of poor outcomes for high risk patients in the market). Price is defined by taking the price for uncomplicated C-section minus the price for uncomplicated natural delivery and averaging over all of the births in each market.

Although the market-level measures throw away a good deal of the variation across providers and the coefficients of interest are generally less precisely estimated, the results are remarkably similar to those discussed above. Better diagnosis (moving from a market at the 25th percentile of the distribution to the 75th) would be associated with an 11.7% decline in C-sections among the low risk and a 3.8% increase in C-sections among the high risk. At the same time, better diagnosis is estimated to decrease the probability of bad outcomes among all risk groups, and to reduce the incidence of neonatal death among the high risk.

A similar improvement in surgical skill relative to skill at natural delivery has little impact on the low risk, but would increase C-sections among the medium and high risk (by 3.2% and 2.2% respectively) and reduce the incidence of bad outcomes in the same two groups. Finally, and as in the physician-level models, an increase in the price gap has the greatest effect on those in the medium risk group, increasing the incidence of C-section by 2.4% and the incidence of bad outcomes by 10.9%.

[10] Appendix Table 2 shows models similar to Tables 4 and 5 except that they use this diagnosis measure for physicians. The results are quite similar to those discussed above.

Overall, the results in this subsection suggest that our results are not driven by the matching of high risk patients to low skilled doctors (which is the only type of selection that could generate a spurious relationship between doctor skill and good outcomes).

5.2 Patient demand

Table 9 shows the results of estimating a demand model. The Alternative Specific Conditional Logit seeks to explain choice among a number of alternatives where the choice set is different for each woman, each choice has characteristics associated with it, and each woman has individual characteristics (such as race and Medicaid coverage) that affect her choice.[11] In order to limit the number of alternatives to a manageable number, we focus on the demand for hospitals, and construct measures of procedural skill, diagnostic skill, and the price gap at the hospital level. Other hospital level characteristics that we include are the distance from the mother's residence to the hospital, and in some specifications, the hospitals' C-section rate, the overall rate of bad outcomes at the hospital, and the fraction of births to Medicaid patients. The individual level characteristics we include are those shown in Table 4 and discussed above as well as the propensity score which determines the woman's a priori risk of C-section.

These results suggest, not surprisingly, that distance to the hospital is of overwhelming importance. Indeed, over sixty percent of the women in our sample deliver at the nearest hospital. However, the estimates all suggest that women value surgical skill, and are more likely to choose a hospital that has good outcomes for high risk patients. In contrast, women are less likely to choose a hospital where providers have good diagnostic skills. Patients

[11] One problem that arises is that about 4% of women deliver in a hospital that is not in their market choice set as we have defined it. Viewed more positively, we correctly identified the market in 96% of the cases. In any case, we drop those women who chose outside the choice set, treating them as if they made those choices for purely idiosyncratic reasons.

are quite insensitive to price, which makes sense given how difficult it is to learn what prices are in most hospitals.

Columns 2 and 3 show that the estimates are not affecting by adding other characteristics of hospitals to the model. For example, patients do not appear to care about the overall C-section rate, or about the rate of bad outcomes. They do however care about the share of Medicaid deliveries, being less likely to choose hospitals with a lot of Medicaid patients, other things being equal.

These results suggest that it is not surprising that bad diagnosis can persist, if patients do not value good diagnosis, do not look at prices or C-section rates, and instead judge hospitals in terms of their prowess dealing with high risk cases surgically.

6 Discussion and Conclusions

We present a model that focuses on diagnostic skill as an element of provider quality that is separate from procedural skill. The model shows that unlike higher surgical skill, which leads to higher use of surgical procedures across the board, better diagnostic skill results in fewer procedures for the low risk, but more procedures for the high risk. That is, better diagnostic skill improves the matching between patients and procedures and leads to better health outcomes.

Taking the model to data on C-sections, the most common surgical procedure performed in the U.S., we show that improving diagnostic skills from the 25th to the 75th percentile of the observed distribution would reduce C-section rates by 11.7% among the low risk, and increase them by 3.8% among the high risk. Since in our application there are many more low risk women than high risk women, improving diagnosis would reduce overall C-section rates without depriving high risk women of necessary care. Moreover, we show that an increase in diagnostic skill would improve health outcomes for both high risk and low risk women, while improvements in surgical skill have much larger effects on high risk women. These results are consistent with the hypothesis that improving diagnosis through methods such as checklists, computer assisted diagnosis, and collaborative decision making could reduce unnecessary procedure use and improve health outcomes.

References

Afendulis, C. C. and D. P. Kessler (2007). Tradeoffs from integrating diagnosis and treatment in markets for health care. *The American Economic Review 97*(3), pp. 1013–1020.

Baicker, K., K. S. Buckles, and A. Chandra (2006, SEP-OCT). Geographic variation in the appropriate use of cesarean delivery. *HEALTH AFFAIRS 25*(5), W355–W367.

Baicker, K., E. S. Fisher, and A. Chandra (2007, MAY-JUN). Malpractice liability costs and the practice of medicine in the medicare program - this analysis suggests that an important association exists between malpractice costs and the use of imaging services in particular. *HEALTH AFFAIRS 26*(3), 841–852.

Baker, G.R. amd MacIntosh-Murray, A., C. Porcellato, K. Dionne, L. amd Stelmacovich, and K. Born (Eds.) (2008). *High Performing Healthcare Systems: Delivering Quality by Design*. Toronto: Longwoods Publishing.

Chandra, A., D. Cutler, and Z. Song (2011). Chapter six - who ordered that? the economics of treatment choices in medical care. In T. G. M. Mark V. Pauly and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume 2 of *Handbook of Health Economics*, pp. 397 – 432. Elsevier.

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

Committee on Identifying and Preventing Medication Errors (2007). *Preventing Medication Errors: Quality Chasm Series*. The National Academies Press.

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

DeGroot, M. H. (1972). *Optimal Statistical Decisions*. New York, NY: McGraw-Hill Book C.

Doi, K. (2007, Jun). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput Med Imaging Graph 31*(4), 198–211.

Dranove, D., S. Ramanarayanan, and A. Sfekas (2011). Does the market punish aggressive experts? evidence from cesarean sections. *B E Journal oOf Economic Analysis & Policy 11*(2).

Dubay, L., R. Kaestner, and T. Waidmann (1999, August). The impact of malpractice fears on cesarean section rates. *Journal of Health Economics 18*(4), 491–522.

Esptein, A. J. and S. Nicholson (2009). The formation and evolution of physician treatment syltes: An application to cesarean sections. *Journal of Health Economics 28*, 1126–1140.

Garber, A. M. and J. Skinner (2008). Is american health care uniquely inefficient? *The Journal of Economic Perspectives 22*(4), pp. 27–50.

Gawande, A. (2009). *The Checklist Manifesto: Getting Things Right*. New York: Picador.

Gruber, J., J. Kim, and D. Mayzlin (1999, AUG). Physician fees and procedure intensity: the case of cesarean delivery. *Journal of Health Economics 18*(4), 473–490.

Gruber, J. and M. Owings (1996). Physician financial incentives and cesarean section delivery. *The RAND Journal of Economics 27*(1), pp. 99–123.

Kessler, D. and M. McClellan (1996). Do doctors practice defensive medicine? *Quarterly Journal of Economics 111*(2), 353–90.

Mello, M. M. and D. M. Studdert (2007). Decontructing negligence: The role of individual and system factors in causing medical injuries. *Georgetown Law Journal 96*, 599–623.

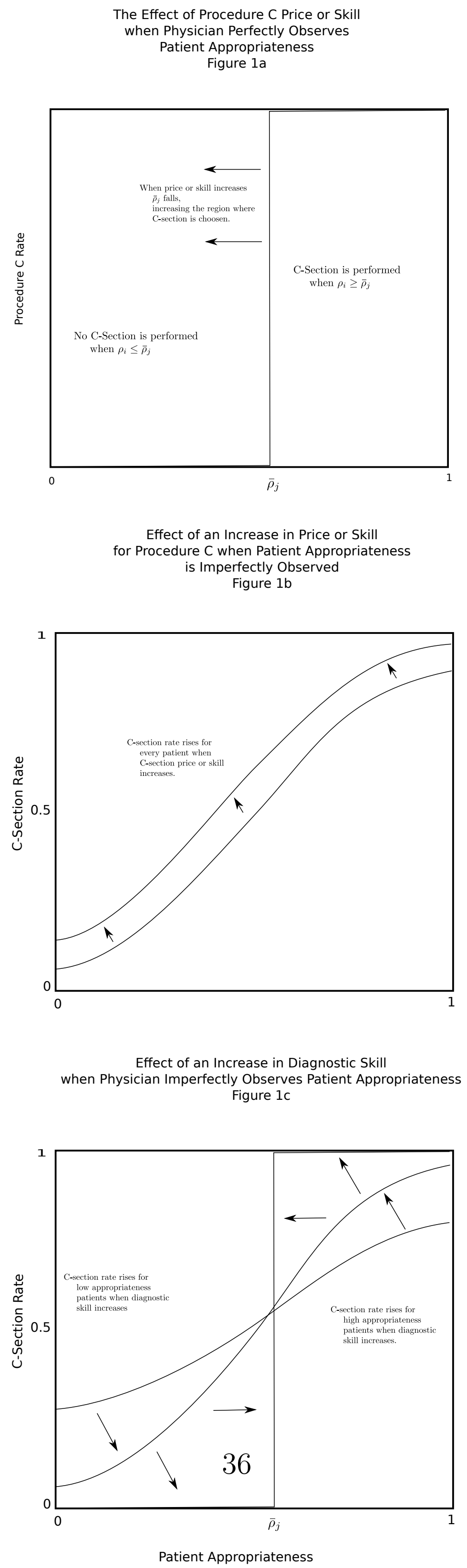


Figure 1: Effect of Patient Appropriateness upon Physician Choice

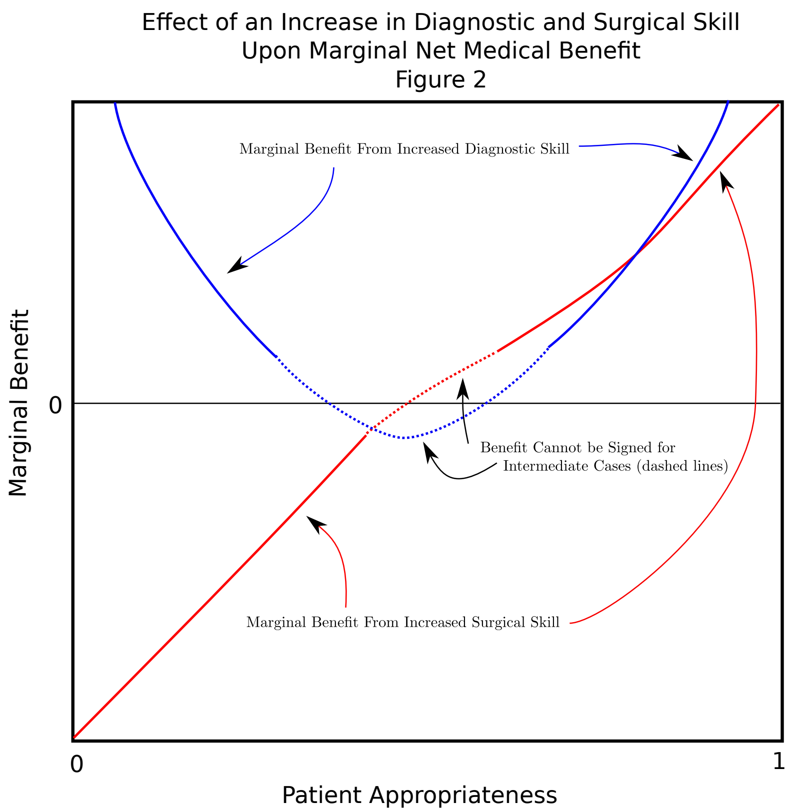
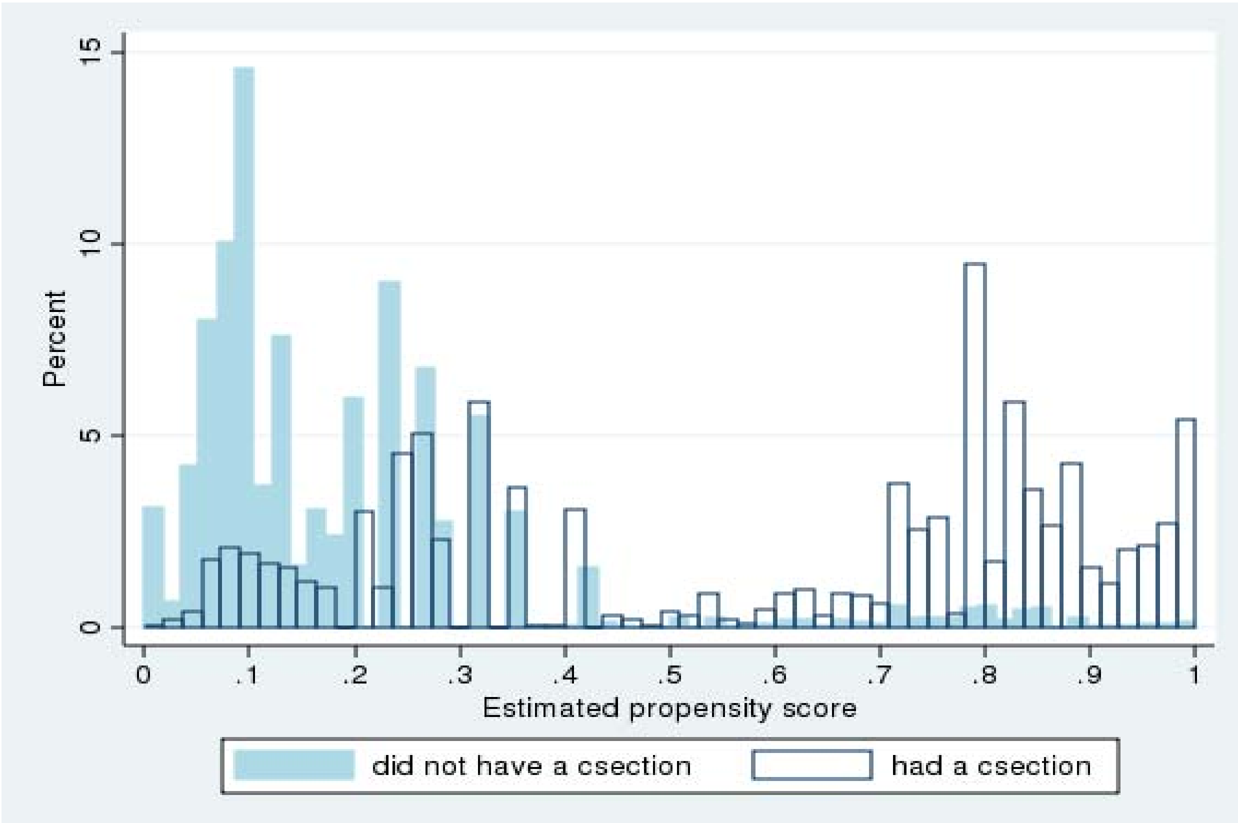


Figure 2: Effect of Skill on Net Medical Benefit

Figure 3: Predicting C-sections Using the Logit Model



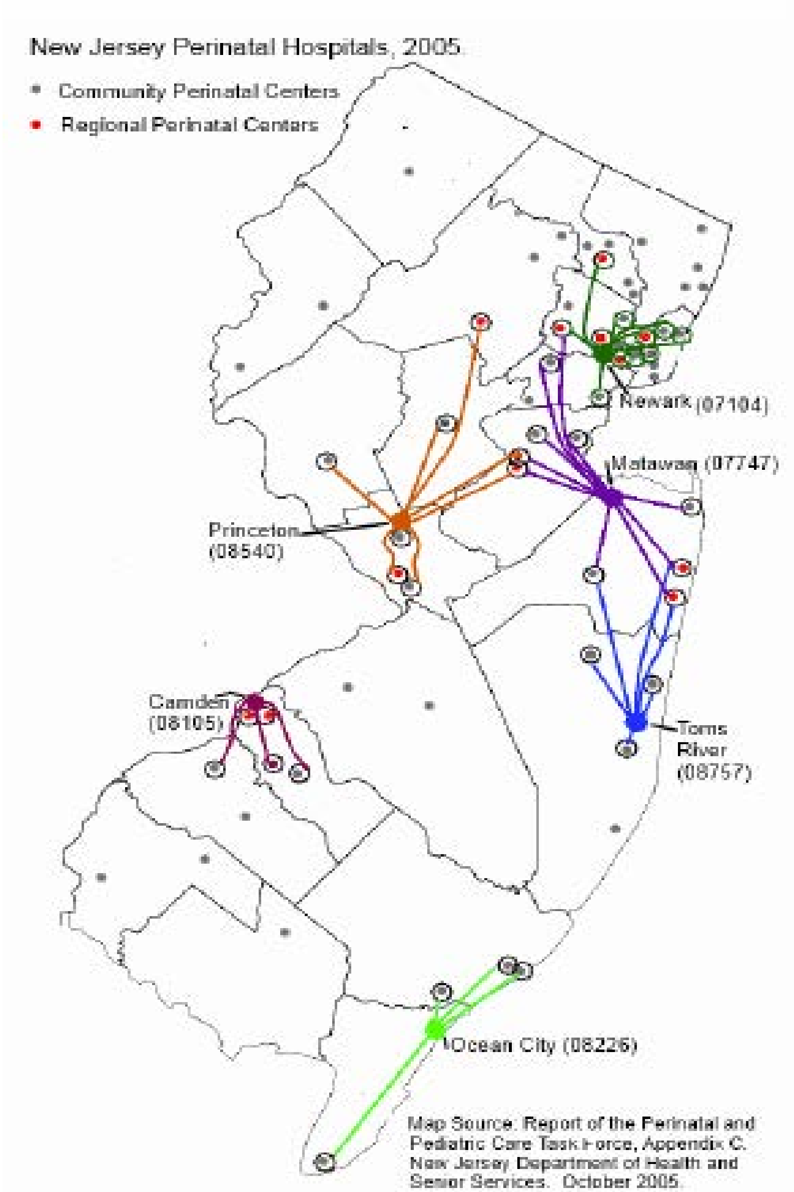
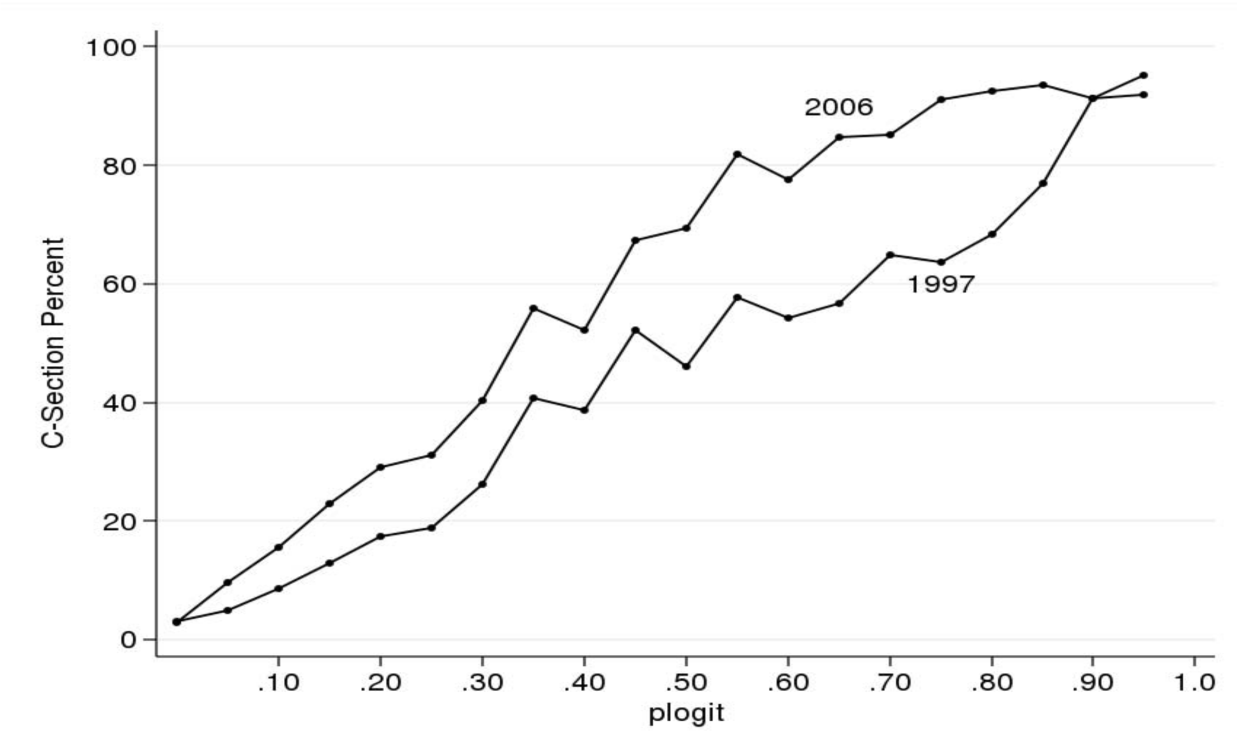


Figure 4: Illustrating the Definition of a Market

Appendix Figure 1: Shift in Probability of C-section Given Medical Risk Over Time



Appendix Figure 2: Shift in Medical Risks over Time

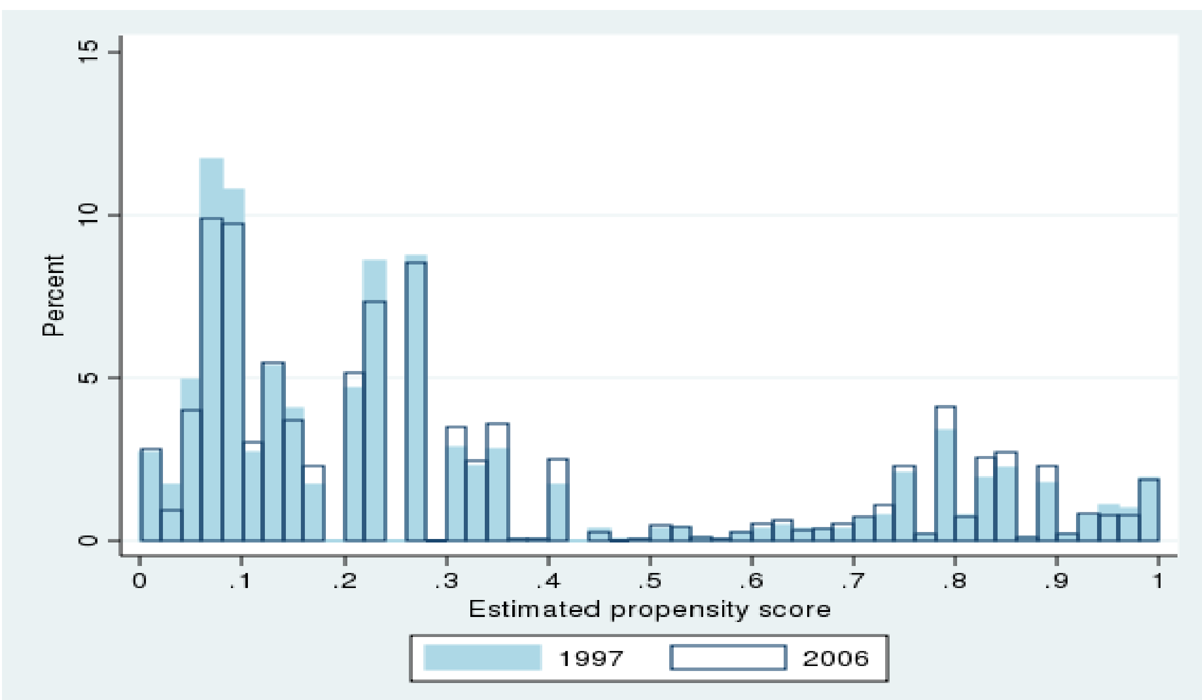


Table 1: A Model of C-section Risk (rho)
(Logistic Regression)

	Coeff.	**S.E.**	**Marginal Effect**
Age<20	-0.337	0.013	-0.075
Age >=25&<30	0.262	0.008	0.058
Age >=30&<35	0.434	0.008	0.096
Age >=35	0.739	0.009	0.164
2nd Birth	-1.347	0.007	-0.298
3rd Birth	-1.645	0.009	-0.364
4th or Higher Birth	-2.140	0.012	-0.474
Previous C-section	3.660	0.008	0.810
Previous Large Infant	0.139	0.029	0.031
Previous Preterm	-0.293	0.025	-0.065
Multiple Birth	2.879	0.014	0.638
Breech	3.353	0.016	0.742
Placenta Previa	3.811	0.054	0.844
Abruptio Placenta	2.048	0.030	0.454
Cord Prolapse	1.761	0.047	0.390
Uterine Bleeding	0.026	0.035	0.006
Eclampsia	1.486	0.096	0.329
Chronic Hypertension	0.745	0.025	0.165
Chronic Lung Conditio	0.064	0.014	0.014
Cardiac Condition	-0.121	0.020	-0.027
Diabetes	0.558	0.011	0.124
Anemia	0.131	0.018	0.029
Hemoglobinopathy	0.116	0.047	0.026
Herpes	0.461	0.024	0.102
Other STD	0.052	0.017	0.012
Hydramnios	0.616	0.018	0.136
Incompetent Cervix	0.043	0.035	0.010
Renal Disease	-0.024	0.031	-0.005
Rh Sensitivity	-0.045	0.040	-0.010
Other Risk Factor	0.276	0.006	0.061
Constant	-1.414	0.007	-0.313
# Observations	1169654		
Pseudo R2	0.32		

Notes: The model also included indicators for missing age, parity, and risk factors.

Table 2: Means by Probability of C-Section

	All	Low	Medium	High
Outcomes				
C-Section Rate	0.331	0.108	0.434	0.887
Any Bad Outcome	0.066	0.047	0.085	0.082
Neonatal death (per 1000)	4.087	2.742	4.803	7.153
Doctor Characteristics				
# Deliveries in Sample	1019	1029	1008	1013
	(650)	(675)	(626)	(627)
Diagnostic Skill	1.033	1.026	1.040	1.039
	(0.183)	(0.185)	(0.180)	(0.181)
Procedural Skill Differential	-0.035	-0.035	-0.034	-0.035
	(0.055)	(0.058)	(0.053)	(0.050)
Price Differential ($1000)	4.756	4.824	4.689	4.697
	(2.678)	(2.723)	(2.630)	(2.642)
Share High Risk	0.126	0.119	0.128	0.141
	(0.044)	(0.042)	(0.043)	(0.048)
Rate of Bad Outcomes	0.066	0.065	0.065	0.068
	(0.042)	(0.040)	(0.041)	(0.044)
Mother & Child Characteristics				
African American	0.157	0.186	0.130	0.128
Hispanic	0.210	0.243	0.178	0.181
Married	0.713	0.645	0.769	0.804
High School Dropout	0.425	0.180	0.079	0.078
Teen mom	0.030	0.056	0.002	0.012
Mom Age 35 or More	0.238	0.226	0.194	0.419
Smoked	0.082	0.092	0.074	0.073
Child Male	0.513	0.514	0.512	0.514
Child First Born	0.397	0.479	0.697	0.273
# of Observations	969140	475270	372239	121631

Notes: The analysis sample excludes birth attendants who were not physicians, and birth attendants who had too few deliveries for a measure of diagnositic skill to be computed. Standard deviations in parentheses.

Table 3: Correlations Between P(C-section) and Doctor Characteristics, Overall and Within Risk Categories

Group= All	**P(C-section)**	**#Deliveries**	**Diagnostic Skill**	**Procedural Skill**	**Price Difference**	**Share High Risk**
P(C-section)	1					
# Deliveries	-0.008	1				
Diagnostic Skill	0.019	0.043	1			
Procedural Skill Diff.	0.000	0.063	0.230	1		
Price Difference	-0.007	0.048	-0.017	0.001	1	
Share High Risk	0.160	-0.027	0.092	-0.020	-0.129	1
Rate of Bad Outcomes	0.030	-0.008	-0.284	-0.403	0.013	0.168
Group= Low Risk						
P(C-section)	1					
# Deliveries	0.002	1				
Diagnostic Skill	0.002	0.049	1			
Procedural Skill Diff.	0.001	0.068	0.235	1		
Price Difference	0.012	0.059	-0.014	0.010	1	
Share High Risk	0.025	0.008	0.141	0.002	-0.137	1
Rate of Bad Outcomes	0.022	-0.099	-0.274	-0.383	0.011	0.135
Group= Medium Risk						
P(C-section)	1					
# Deliveries	0.009	1				
Diagnostic Skill	-0.020	0.036	1			
Procedural Skill Diff.	-0.016	0.057	0.227	1		
Price Difference	0.041	0.035	-0.021	-0.009	1	
Share High Risk	0.038	-0.035	0.051	-0.032	-0.121	1
Rate of Bad Outcomes	0.036	-0.093	-0.295	-0.409	0.019	0.174
Group= High Risk						
P(C-section)	1					
# Deliveries	-0.021	1				
Diagnostic Skill	-0.014	0.046	1			
Procedural Skill Diff.	-0.054	0.059	0.215	1		
Price Difference	-0.041	0.033	-0.010	-0.006	1	
Share High Risk	0.115	-0.129	0.010	-0.082	-0.111	1
Rate of Bad Outcomes	0.06	-0.09	-0.298	-0.476	0.007	0.252

Note: All of the correlations are statistically significant at the 95% level of confidence.

Table 4: Effect of Doctor Variables on Probability of C-Section

		C-section	
Medical Risk:	**Low**	**Medium**	**High**
Diagnostic Skill	-0.059	0.062	0.190
	(0.009)	(0.014)	(0.009)
Procedural Skill Difference	0.094	0.131	0.135
	(0.022)	(0.043)	(0.033)
Price Differential (x 100)	0.335	0.545	-0.021
	(0.058)	(0.096)	(0.054)
C-section Risk	0.860	0.704	0.804
	(0.033)	(0.022)	(0.020)
African-American	0.053	0.066	0.027
	(0.003)	(0.006)	(0.004)
Hispanic	0.024	0.056	0.033
	(0.003)	(0.005)	(0.004)
Less than High School	0.023	0.033	0.017
	(0.002)	(0.004)	(0.005)
High School	0.028	0.042	0.024
	(0.002)	(0.003)	(0.003)
Some College	0.014	0.019	0.007
	(0.001)	(0.002)	(0.003)
Married	-0.010	-0.009	0.004
	(0.002)	(0.003)	(0.003)
Medicaid	0.007	0.0004	0.013
	(0.002)	(0.004)	(0.003)
Teen Mom	-0.024	-0.048	0.019
	(0.003)	(0.019)	(0.010)
Mother 25-34	0.030	0.030	0.008
	(0.002)	(0.004)	(0.004)
Mother 35+	0.045	0.070	0.011
	(0.003)	(0.005)	(0.004)
Mother Smoked	0.012	0.006	0.001
	(0.002)	(0.003)	(0.004)
Child Male	0.019	0.034	0.005
	(0.001)	(0.002)	(0.002)
Child 2nd Born	-0.045	0.185	-0.001
	(0.004)	(0.011)	(0.003)
Child 3rd Born	-0.049	0.136	-0.037
	(0.005)	(0.010)	(0.004)
Child 4th Born or Higher	-0.040	0.064	-0.043
	(0.006)	(0.010)	(0.005)
R-squared	0.041	0.213	0.063
# Observations	475270	372239	121631

Notes: Standard errors clustered by physician. Regressions also included month and year of birth indicators, and indicators for missing educaton, marital status, Medicaid, smoking, prices, and parity.

Table 5: Effect of Doctor Variables on Probability of Negative Outcomes

	Any Bad Outcome			Neonatal Death		
Medical Risk:	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Diagnostic Skill	-0.029	-0.048	-0.028	-0.004	-0.011	-0.015
	(0.005)	(0.008)	(0.008)	(0.001)	(0.001)	(0.002)
Procedural Skill Difference	-0.070	-0.314	-1.073	-0.004	-0.021	-0.084
	(0.024)	(0.039)	(0.035)	(0.002)	(0.005)	(0.010)
Price Differential (x 100)	0.066	0.022	0.054	0.013	0.022	-0.007
	(0.042)	(0.055)	(0.050)	(0.004)	(0.007)	(0.011)
C-section Risk	0.345	-0.089	0.356	0.014	-0.036	0.0340
	(0.020)	(0.014)	(0.019)	(0.005)	(0.004)	(0.007)
R-squared	0.011	0.009	0.058	0.004	0.009	0.016
# Observations	475270	372239	121631	475270	372239	121631

Notes: Standard errors are clustered on the physician and shown in parentheses.
Regressions also included all of the variables listed in Table 4.

Table 6: Effect of Doctor Variables Including Share High Risk

	C-Section			Any Bad Outcome			Neonatal Death		
Medical Risk:	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Diagnostic Skill	-0.067	0.058	0.190	-0.032	-0.049	-0.029	-0.004	-0.011	-0.015
	(0.008)	(0.013)	(0.009)	(0.005)	(0.007)	(0.007)	(0.001)	(0.001)	(0.002)
Procedural Skill Difference	0.103	0.152	0.156	-0.066	-0.306	-1.06	-0.003	-0.020	-0.083
	(0.021)	(0.042)	(0.033)	(0.023)	(0.037)	(0.034)	(0.002)	(0.005)	(0.010)
Price Differential (x 100)	0.368	0.607	0.008	0.081	0.044	0.067	0.013	0.025	-0.007
	(0.057)	(0.094)	(0.054)	(0.041)	(0.054)	(0.049)	(0.004)	(0.007)	(0.011)
C-section Risk	0.822	0.692	0.782	0.327	-0.093	0.347	0.013	-0.036	0.034
	(0.032)	(0.021)	(0.020)	(0.020)	(0.014)	(0.019)	(0.005)	(0.004)	(0.007)
Share High Risk in Practice	0.332	0.575	0.257	0.151	0.203	0.117	0.007	0.024	0.006
	(0.040)	(0.074)	(0.036)	(0.024)	(0.033)	(0.029)	(0.003)	(0.007)	(0.008)
R-squared	0.043	0.215	0.065	0.012	0.010	0.059	0.004	0.009	0.016
# Observations	475270	372239	121631	475270	372239	121631	475270	372239	121631

Notes: Standard errors are clustered on the physician and shown in parentheses.
Regressions also included all of the variables listed in Table 4.

Table 7: Effect of Doctor Variables Excluding Planned C-Sections from Sample

	C-Section			Any Bad Outcome			Neonatal Death		
Medical Risk:	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Diagnostic Skill	-0.033	0.006	0.333	-0.026	-0.045	-0.043	-0.003	-0.009	-0.022
	(0.006)	(0.013)	(0.019)	(0.005)	(0.008)	(0.012)	(0.001)	(0.001)	(0.004)
Procedural Skill Difference	0.071	0.13	0.119	-0.066	-0.237	-1.22	-0.004	-0.015	-0.136
	(0.016)	(0.035)	(0.060)	(0.024)	(0.037)	(0.047)	.002)	(0.006)	(0.021)
Price Differential (x 100)	0.189	0.37	-0.225	0.062	(0.018)	(0.057)	0.011	0.030	0.002
	(0.043)	(0.095)	(0.128)	(0.042)	(0.057)	(0.094)	(0.003)	(0.008)	(0.028)
C-section Risk	0.418	0.357	2.481	0.320	-0.005	0.396	0.006	-0.035	0.072
	(0.025)	(0.019)	(0.052)	(0.021)	(0.015)	(0.037)	(0.005)	(0.005)	(0.018)
R-squared	0.044	0.033	0.208	0.011	0.005	0.057	0.004	0.014	0.026
# Observations	455834	282265	43119	455834	282265	43119	455834	282265	43119

Notes: Standard errors are clustered on the physician and shown in parentheses.
Regressions also included all of the variables listed in Table 4.

Table 8: Effect of Market Level Variables

	C-Section			Any Bad Outcome			Neonatal Death		
Medical Risk:	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Market Diagnostic Skill	-0.203	0.022	0.550	-0.073	-0.055	-0.099	0.000	-0.003	-0.031
	(0.034)	(0.051)	(0.059)	(0.020)	(0.027)	(0.047)	(0.004)	(0.007)	(0.014)
Market Procedural Skill Differe	0.071	0.279	0.390	0.011	-0.103	-0.614	0.005	-0.008	-0.046
	(0.045)	(0.074)	(0.095)	(0.034)	(0.052)	(0.092)	(0.009)	(0.012)	(0.028)
Market Price Differential (x 10	0.142	0.395	-0.143	-0.010	0.355	0.202	-0.001	0.035	0.004
	(0.114)	(0.142)	(0.150)	(0.054)	(0.075)	(0.119)	(0.001)	(0.018)	(0.004)
C-section Risk	0.883	0.711	0.793	0.338	-0.094	0.392	0.014	-0.036	0.039
	(0.033)	(0.018)	(0.019)	(0.018)	(0.012)	(0.018)	(0.005)	(0.004)	(0.007)
R-squared	0.046	0.218	0.062	0.018	0.013	0.036	0.006	0.011	0.017
# Observations	475270	372239	121631	475270	372239	121631	475270	372239	121631

Notes: Standard errors are clustered on the zip code and shown in parentheses.
Regressions also included all of the variables listed in Table 4 as well as zip code fixed effects.

Table 9: Models of Hospital Demand

	Coeff.	Coeff.	Coeff.
Distance from Residence	-0.256	-0.256	-0.256
	(0.000)	(0.000)	(0.000)
Procedure Skill Differential	0.703	0.596	0.590
	(0.048)	(0.048)	(0.050)
Diagnostic Skill	-0.245	-0.260	-0.261
	(0.034)	(0.034)	(0.035)
Price Difference (CS-N)	0.0017	0.0013	0.0013
	(0.0009)	(0.0009)	(0.0009)
C-section Rate	..	0.005	0.007
		(0.053)	(0.053)
Medicaid Rate	..	-0.437	-0.435
		(0.022)	(0.023)
Rate of Bad Outcomes	..	..	-0.032
			(0.072)

Notes: Standard errors in parentheses. Estimated using Alternative Specific Conditional Logit. Patient characteristics (maternal age, race, education, Medicaid coverage) are allowed to have separate effects on each hospital choice.

Appendix Table 1: Effect of Doctor Variables Including Share High Risk

	C-Section			Any Bad Outcome			Neonatal Death		
Medical Risk:	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Diagnostic Skill	-0.065	0.054	0.187	-0.031	-0.050	-0.030	-0.004	-0.011	-0.015
	(0.009)	(0.013)	(0.009)	(0.005)	(0.008)	(0.007)	(0.001)	(0.001)	(0.002)
Procedural Skill Difference	0.099	0.149	0.144	-0.065	-0.306	-1.07	-0.003	-0.020	-0.083
	(0.021)	(0.041)	(0.032)	(0.024)	(0.037)	(0.034)	(0.002)	(0.005)	(0.010)
Price Differential (x 100)	0.371	0.611	0.010	0.082	0.045	0.067	0.013	0.024	-0.008
	(0.057)	(0.094)	(0.053)	(0.041)	(0.054)	(0.049)	(0.004)	(0.007)	(0.011)
C-section Risk	0.830	0.689	0.791	0.329	-0.094	0.351	0.013	-0.036	0.0340
	(0.032)	(0.022)	(0.020)	(0.019)	(0.014)	(0.019)	(0.005)	(0.004)	(0.007)
Share High Risk Lowest	-0.031	-0.056	-0.032	-0.015	-0.019	-0.011	-0.001	-0.002	-0.001
Quartile	(0.005)	(0.008)	(0.005)	(0.003)	(0.004)	(0.003)	(0.0003)	(0.0006)	(0.001)
Share High Risk Second	-0.018	-0.036	-0.013	-0.013	-0.016	-0.010	-0.001	-0.001	0.000
Quartile	(0.004)	(0.007)	(0.004)	(0.003)	(0.004)	(0.003)	(0.0003)	(0.0005)	(0.001)
Share High Risk Third	-0.010	-0.015	-0.010	-0.008	-0.012	-0.006	-0.001	-0.002	-0.001
Quartile	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.003)	(0.0003)	(0.0005)	(0.001)
R-squared	0.043	0.215	0.064	0.0118	0.010	0.059	0.004	0.009	0.016
# Observations	475270	372239	121631	475270	372239	121631	475270	372239	121631

Notes: Standard errors are clustered on the physician and shown in parentheses.
Regressions also included all of the variables listed in Table 4.