Doctor's Diagnostic Skill, Procedural Skill and Unnecessary C-Sections

Janet Currie

Princeton University, IZA, NBER

W. Bentley MacLeod

Columbia University, IZA, NBER

Dec. 1, 2012

Revised Sept. 17, 2013

Abstract

This paper develops and applies a model of in which doctors have two dimensions of skill: diagnostic skill and skill performing procedures. Higher procedural skill increase the use of intensive procedures across the board, while better diagnostic skill results in fewer intensive procedures for the low risk, but more for the high risk. Deriving empirical analogues to our theoretical measures for the case of C-section, we show that improving diagnostic skill would reduce C-section rates by 15.8% among the lowest risk, and increase them by 4.7% among the high risk while improving outcomes among all women.

*We thank Samantha Heep and Dawn Koffman for excellent research assistance, and Amitabh Chandra, Jonathan Gruber, Amy Finkelstein, Kate Ho, Robin Lee, Jonathan Skinner and seminar participants at Princeton, Georgetown University, Havard Medical School, Kyoto University, NYU, the Japanese National Institute of Population and Social Security Research, Warwick University, University College London, the London School of Economics, the Paris School of Economics, the NBER Summer Institute, 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.

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 explores the idea that variations in treatment arise because some physicians are better than others at using the available information to make a decision about treatment, a capability we dub "diagnostic skill." Most previous analyses of physician decision making have focused on a single dimension of physician skill, viz. physician skill in performing procedures, and have ignored diagnosis. Instead, in one of the few attempts to go beyond a uni-dimensional model of skill, we develop a model in which physician skill has two dimensions: Physicians 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, which in turn will affect the doctor's decision about treatment. Building on learning models such as Farber and Gibbons (1996) and Altonji and Pierret (2001), 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. Our work also draws inspiration from new research in management (e.g. Bloom and Van Reenen (2010)) which suggests that

[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.

decision making skill is an important ingredient of performance.

Although it has been neglected in the health economics literature, diagnostic skill has become increasingly important because of the growing complexity of medical care and the sheer number of different treatment options available. For example, in a world in which there was little that could be done for cancer patients, it did not matter if physicians choose the right treatment; 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]). An important reason to try to measure diagnostic decision making skill 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 of quantifying diagnostic decision making skill and relating it to outcomes, C-section, which is the most common surgical procedure in the U.S., is ideal: 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 first, that it is theoretically and empirically possible to distinguish between procedural and diagnostic skill. Second, we develop meaningful empirical proxies for these concepts. Third we show that the predictions of the model are borne out in the

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

data: Improvements in diagnostic decision making skill increase the incidence of C-sections for high risk women, but reduce C-sections for low risk women. Since low risk women outnumber high risk ones, improving diagnosis reduces overall C-section rates. This reduction does not result from across-the-board cuts in C-section rates. Instead, we estimate that moving a woman from a provider at the 25th to the 75th percentile of the distribution of diagnostic skill would reduce the probability of C-section among the lowest risk women by 15.8%, but would increase the probability of C-section among high-risk women by 4.7%. By way of comparison, increasing providers' procedural skill performing C-sections by a comparable amount would increase C-section rates by about 3.7% among low risk women, but by only .5% among high risk women. Moreover, since most low risk women are 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 17.0% among the lowest risk women, and by 8.3% among high risk women. In contrast, improving surgical skill per se mainly benefits high risk women and may even have negative effects on the lowest risk women by encouraging unnecessary procedure use.

By highlighting the importance of diagnostic skill in addition to procedural skill, and suggesting empirical analogs of these empirical concepts, our paper takes a first step towards measuring and improving diagnostic decision making. 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 (although Grant (2009) argues that the effect is smaller than they had estimated). 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: Epstein 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 this model, the choice of procedure depends only on the technical procedural skill of the physician. Taking our cue from the literature on management effectiveness (Bloom and Van Reenan, 2010) and variations in business productivity (Syverson, 2011, Finkelstein and Syverson, 2013), we build on Chandra-Staiger by exploring the hypothesis that part of the variation in treatment choice is due to variation in decision making skills.

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 decision making 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. 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. We also show that our measures of these two dimensions of skill are positively correlated, as one might expect, but that it is possible to distinguish them since the correlation is a modest .259.

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 begin with the standard Roy model of physician decision making to understand physician diagnosis, and then add to this model Bayesian learning.[4] In our data we observe patient characteristics, the procedure chosen, and various measures of medical outcomes. The goal is to understand how variations in physician skill affect

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

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 Physician Behavior

Suppose that the physician chooses the best action possible given her information, costs, and patient preferences. The procedure that doctor j chooses for patient i, is 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. When deciding upon a procedure the physician evaluates the underlying condition of the patient to produce two latent variable H_i^N and H_i^C which are the outcomes if procedure N or C are performed. We assume that physicians care both about patient outcomes and about the fees they can charge. In turn, the patient's outcome depends on the underlying condition of the patient, the procedure chosen, and on the physician's skill. Hence, the utility payoff of the physician is:

$$U_{ij}\left(T\right) = (H_i^T)^{\alpha^H} (S_j^T)^{\alpha^S} M_j^T \left(P^T\right)^{\alpha^M}, \tag{1}$$

where S_j^T is the skill of physician j 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. The elasticities of the interaction terms are given by $\alpha^k, k = H, S, M$. Taking logs yields:

$$\begin{aligned} u_{ij}\left(T\right) &= log(U_{ij}\left(T\right)) \\ &= \alpha^H log\left(H_i^T\right) + \alpha^S log\left(S_j^T\right) + \alpha^M log\left(M_j^T\left(P^T\right)\right) \end{aligned}$$

Without loss of generality, we can redefine the health outcome, skill, and pecuniary returns by

$$h_i^T = \alpha^H log\left(H_i^T\right),$$
$$s_j^T = \alpha^S log\left(S_j^T\right)$$
$$m_j^T\left(P^T\right) = \alpha^M log\left(M_j^T\left(P^T\right)\right).$$

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

$$MB_{ij}^T = h_i + s_j^T.$$

And the net medical benefit is given by:

$$NMB\left(h_i, s_j^C, s_j^N\right) = MB_{ij}^C - MB_{ij}^N \quad (2)$$
$$= h_i + s_j^C - s_j^N \quad (3)$$

where $h_i = h_i^C - h_i^N$.

The physician cannot directly observe patient condition h_i, but rather estimates the condition of the patient with the available information I_{ij}. We will show presently that we can write:

$$E\{h_i|I_{ij}\} = \bar{h}_{ij} + \epsilon_{ij},$$

where ϵ_{ij} is normally distributed with mean zero and variance σ_j. The physician carries out a C-section if and only if the net medical benefit plus the pecuniary benefit is positive:

$$E\left\{NMB\left(h_i, s_j^C, s_j^N\right) + m_j\left(\Delta P\right)|I_{ij}\right\} \geq 0. \tag{4}$$

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

$$\rho_{ij} = F\left(\left(s_j^C - s_j^N + m_j\left(\Delta P\right) - \bar{s}_{ij}\right)/\sigma_{ij}\right). \tag{5}$$

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. *Taking physician information as given, the rate of use of procedure C increases with physician skill doing procedure C (s_j^C) and with the difference in price between procedure C and procedure N. The use of procedure C falls with an increase in skill doing 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.

3.3 Understanding Diagnosis

Diagnosis means the ability to reliably transform observed symptoms into an assessment of patient condition. Accurate diagnosis is essential to appropriate treatment, though treatment will also depend 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.

If all doctors learn and evaluate information in the same way, then, with sufficient controls for patient characterstics, conditional upon patient condition X_i observed decisions should not statistically vary between physicians. We postulate that diagnosis can be viewed as a *learning process.* By this we mean that when dealing with child birth each physician has a baseline treatment style (as in Epstein and Nicholson (2009)). When the physician observes the patient's condition she learns things that may lead her to change her beliefs regarding the best course of action. We model physician learning as a one step Bayesian updating process. The physician is assumed to have some prior beliefs that correspond to her treatment style. She then observes X_i which updates her beliefs regarding patient condition, and decides to perform a C-section or not.

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:

$$\rho_i^r = F\left(\beta^r X_i\right), \tag{6}$$

where F is a logistic distribution, and $\rho_i^r \in [0,1]$ is the predicted probability of procedure C. Let $h_i^r = \beta^r X_i \in \Re$ be the corresponding index that varies over the real line. We show that ρ_i^r is a physician-independent index 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, $h_i \in \Re$, 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 $h_i \geq 0$. Thus, we can interpret $h_i^r = \beta^r X_i$ 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. This index depends only on the patient's medical condition and is independent of physician characteristics and other patient characteristics (such as race and insurance coverage).

We already know that different physicians often make different decisions with the same data regarding a women's condition which may be in part because they differ in the way that they process information.

We formally capture this effect by supposing that h_i^r , given by equation (6), is an unbiased signal of the net benefit with variance σ_r^2. The physician is assumed to observe:

$$h_{ij} = h_i^r + \epsilon_{ij}/D_j, \tag{7}$$

where the variance of ϵ_{ij} is a standard normal distribution 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 which doctors observe additional data that we do not observe is discussed in section 3.4.2 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 h_j^0 be the mean and σ_{ij}^{02} be the variance of these beliefs. Let $D_j^0 = \frac{1}{\sigma_j^0}$ be a measure of how strongly a physician holds his or her pre-existing beliefs.

[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.

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

$$\begin{aligned} E\left\{h_i|h_{ij},h_j^0\right\} &= \pi h_{ji} + (1-\pi)\, h_j^0 \\ &= \pi h_i^r + (1-\pi)\, h_j^0 + \frac{\pi}{D_j}\epsilon_{ij}, \end{aligned} \tag{8}$$

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 5. If the physician can observe h_i^r directly, then D_j is zero and diagnosis is not an issue. Procedure C is choosen if and only if:

$$h_i^r \geq s_j^N - s_j^C - m\,(\Delta P)\,. \tag{9}$$

This rule is illustrated in Figure 1a where $\bar{\rho}_j = F\left(s_j^N - s_j^C - m\,(\Delta P)\right)$. 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 improvements in surgical skill have their greatest impact on the use of C-section among marginal patients.

3.4 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:

$$E\left\{h_i|I_{ij}\right\} + s_j^C - s_j^N + m_j\left(\Delta P\right) \geq 0. \tag{10}$$

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 (8) to get:

$$\pi h_i^r + (1-\pi)\, h_j^0 + s_j^C - s_j^N + m_j\left(\Delta P\right) \geq \frac{\pi}{D_j}\epsilon_{ij}.$$

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

$$D_j\left(h_i^r - a_j\right) \geq \epsilon_{ij} \tag{11}$$

Where, given prices,

$$a_j = -\frac{(1-\pi)\, h_j^0 + s_j^C - s_j^N + m_j\left(\Delta P\right)}{\pi} \tag{12}$$

is a physician specific constant. Let the probability that a patient i with observed condition h_i^r who is treated by physician j receives procedure C be denoted by ρ_{ij}. Since the the error term ϵ_{ij} is a standard normal distribution from (11) we have:

$$\rho_{ij} = F\left(D_j\left(h_i^r - a_j\right)\right). \tag{13}$$

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.

Equation 13 formalizes the sense in which our model incorporates two dimensions of doctor's skill rather than one dimension. In the standard Roy model, as used for example by Chandra and Staiger (2007) and Epstein and Nicholson (2009), only the constant term a_j varies across physicians (or across regions). Here, in addition to this doctor specific constant, there is a slope term, D_j, which we interpret as a measure of diagnostic skill. 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. In our model, an increase in C-section skill leads to more C-sections:

$$\frac{1}{f\left(D_j\left(h_i^r - a_j\right)\right)} \frac{\partial \rho_{ij}}{\partial s_j^C} = \frac{D_j}{\pi} > 0. \tag{14}$$

However,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 13 with respect to diagnostic skill we get:

$$\frac{1}{f\left(D_j\left(h_i^r - a_j\right)\right)} \frac{\partial \rho_{ij}}{\partial D_j} = h_i^r - b_j, \tag{15}$$

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

$$b_j = a_j + D_j \frac{\partial a_j}{\partial D_j}. \tag{16}$$

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

$$D_j \frac{\partial a_j}{\partial D_j} = \left\{ \left((1-\pi)^2 + \pi \right) h_j^0 + s_j^C - s_j^N + m_j \left(\Delta P\right) \right\} \frac{2\left(1-\pi\right)}{\pi}. \tag{17}$$

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

$$h_j^0 \geq -\left(\frac{s_j^C - s_j^N + m_j\left(\Delta P\right)}{\left(\left(1-\pi\right)^2 + \pi\right)}\right). \tag{18}$$

However, given that the value of b_j does not vary with the condition of the patient and h_i^r can take any real valued expression 15 implies:

Proposition 2. *The probability that the physician uses procedure C increases with diagnostic skill if and only if patient condition is above a fixed, physician specific, threshold ($h_i^r > b_j$).*

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.

3.4.1 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:

$$h_i^r + s_j^C - s_j^N.$$

The physician observes a signal h_{ij} and decides on the procedure following rule 11. We can write the net medical benefit as function of observed medical appropriateness for procedure C as:

$$\begin{aligned} NMB_j\left(h_i^r\right) &= \rho_{ij}\left(h_i^r + s_j^C - s_j^N\right) \\ &- \left(1-\rho_{ij}\right)\left(h_i^r + s_j^C - s_j^N\right) \\ &= \left(2\rho_{ij} - 1\right)\left(h_i^r + s_j^C - s_j^N\right) \qquad (19) \end{aligned}$$

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

$$\frac{\partial NMP_j}{\partial D_j} = 2\frac{\partial \rho_{ij}}{\partial D_j}\left(h_i^r + s_j^C - s_j^N\right)$$

Recall that h_i^r takes values over the whole real line. When h_i^r is sufficiently large then $\frac{\partial \rho_{ij}(h_i^r)}{\partial D_j} > 0$, and the term $(h_i^r + s_j^C - s_j^N)$ is positive; hence diagnositic skill has a positive effect upon outcomes. Similarly, when h_i^r is sufficiently small, $\frac{\partial \rho_{ij}(h_i^r)}{\partial D_j} < 0$, and the term $(h_i^r + s_j^C - s_j^N)$ 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:

$$\frac{\partial NMP_j}{\partial s_j^C} = 2\frac{\partial \rho_{ij}}{\partial s_j^C}\left(h_i^r + s_j^C - s_j^N\right) + \left(2\rho_{ij} - 1\right).$$

Better surgical skill (relative to natural delivery) always increases the number of C-sections. For high risk patients, $h_i^r > max\left\{a_j, s_j^N - s_j^C\right\}$, so both $\left(h_i^r + s_j^C - s_j^N\right)$ and $(2\rho_{ij} - 1)$ are positive. Hence, the effect of skill is positive. We have a negative sign when $h_i^r < min\left\{a_j, s_j^N - s_j^C\right\}$, and hence skill has a negative effect on net benefits for the lowest 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 mis-match 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. The effect of price is quite straightforword and given by:

$$\frac{\partial NMP_j}{\partial \Delta P_j} = 2\frac{\partial \rho_{ij}}{\partial \Delta P_j}\left(h_i^r + s_j^C - s_j^N\right).$$

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

3.4.2 Alternative Information Structures

We have assumed that not all doctors interpret patient conditions h_i^r in the same way. That is, different doctors have different values of D_j. An alternative assumption is that all doctors interpret h_i^r in the same way but some doctors observe additional information. In this case, variations in decisions would be due to the additional information that is collected rather than to physicians processing the same information in different ways. In this alternative scenario, a Bayesian decision maker would put less weight upon h_i^r as she acquired additional information. This in turn would imply that sensitivity to h_i^r would *decrease* with improvements in a physician's diagnostic skills. Recall that in our model a sensitivity toh_i^r is captured by the slope term, D_j.

Hence, this alternative scenario implies that decreases (rather than increases) in D_j would improve outcomes, a result that is derived more formally in the appendix. That is, suppose that we judged a woman to be at very low risk of C-section, but the doctor was able to discern private information that indicated a C-section was necessary, or conversely, suppose we judged a woman to be very high risk, but the doctor was able to ascertain that a normal delivery was safe. In these cases, the relationship between our measure of the woman's health status and C-section risk would be flatter for the skilled diagnositican than for a less skilled colleague, and this flatter relationship would be associated with better outcomes.

As we show below, we find exactly the opposite result. That is, a stronger relationship between our measure of patient risk and the doctor's propensity to do a C-section is predictive of better outcomes. This result suggests that many doctors do not use the information contained in our measures of patient condition,h_i^r, efficiently. Another way to think about this issue is to reflect on the fact that measures of patient risk that we estimate reflect the combined experience of all physicians 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 level of patient risk from the underlying information about patient condition.

4 Data and Methods

The data for this project come from approximately a million Electronic Birth Certificates, (EBC) spanning 1997 to 2006, from the state of New Jersey. These records have several important features. First, in addition to information about the method of delivery including whether a C-section was planned or not, 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, 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.

Third, 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 our analysis, we focus on doctors and exclude midwives since only doctors can perform C-sections.

Finally, the data includes demographic information about the mother 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.

The inclusion of these variables may help us to control for variations in demand for C-sections by different demographic groups.

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 use is shown in column 1 of Table 1. Table 1 shows that the model predicts well, with a pseudo R-squared of almost .32. One issue with this model is that it reflects actual practice, but not necessarily best practice. 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 several alternative models and found that the correlation between the ranking produced by our model, and the ranking produced by the alternatives is above .95. These alternatives included a model with 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. Estimates of the latter model are also shown in Table 1. One can see that the estimated coefficients for these "good doctors" are similar to those for all doctors suggesting that there is not a lot of controversy about which women are the best candidates for C-section. Rather, the controversy about C-section can be interpreted as a matter of where the cutoff for C-section should occur.

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 .06 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 "very low risk" and the "high risk" respectively, and consider those with values of ρ_i^r between .06 and .2 and between .2 and .8 as "low risk" and "medium risk," respectively. Of the women deemed high risk, 89% received a C-section, while among the women deemed very low risk only 6% 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%. Appendix Figure 2 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).

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 h_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 h_i^r. The coefficient on h_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 very 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, the vast majority of high risk women get C-sections and most very low risk women do not. At the same time, because the high risk and very 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 we have defined it that way so that 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 list charges for every discharge. For each hospital and year, 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 form:

$$Outcome_{ijt} = f(DiagSkill_j, s_j^C - s_j^N, \Delta P_{jt}, Z_{it}, month, year), \tag{20}$$

where $Outcome_{ijt} \in \{0, 1\}$, where 0 is a Natural delivery (or good birth outcome) and 1 is a C-Section (or bad birth outcome), i indexes the patient, j indexes the doctor, and t indexes the year. The vector Z_{it} includes maternal age (less than 20, 20-24, 25-29, 30-34, 35 and over), education (less than 12, 12, 13-15, 16 or more), marital status, race/ethnicity (African-American, Hispanic), and whether the birth was covered by Medicaid, as well as the child's gender and indicators for birth order. 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 while we used all births to calculate the probability of C-section, in the rest of the paper we exclude births that were not attended by a doctor, as well

[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.

as those for whom we cannot calculate our measure of diagnostic skill (because there are too few births per provider).[7] These exclusions leave us with approximately 1,000 providers, who together deliver the vast majority of the babies in New Jersey over the sample period. The first panel shows how the outcome variables vary across the four risk groups. As expected, higher risk women have more C-sections, a higher risk of a bad outcome, and higher neonatal death rates compared to lower risk women.

The second panel explores the characteristics of doctors and provides some initial evidence with regard to an important question: The extent to which higher risk patients see doctors with particular characteristics. Table 2 suggests that the doctors who treat low, medium, and high risk patients are remarkably similar in terms of number of deliveries in the sample, diagnostic skill, procedural skill, and price differentials. There is however a clear gradient in the share of high risk patients in the practice, with high risk patients being more likely to see doctors who are relatively more specialized in high risk patients. The lowest risk patients appear to see doctors who are slightly less skilled and who see more patients than average. Perhaps surprisingly, there is little difference in the fraction of a doctor's patients who have had 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 fractions of patients with bad outcomes in their practices.

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 the lowest risk women are disproportionately minority women who have already had at least one birth, whereas women at high risk for C-section tend to be older, married women.

The main empirical difficulty involved in estimating (20) is that women choose their doctors. If women with high risk pregnancies choose better doctors, then the

[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.

estimated effect of doctor skill on birth outcomes will be biased towards zero. Table 2 suggests that there is some evidence of this type of selection, particularly for the lowest risk group. We do not see any evidence that high risk women go to the least skilled doctors, which would lead estimates of (20) to overstate the beneficial effect of skill on birth outcomes. Although there is no perfect solution to the problem of doctor selection, we address it in several ways.

First, we examine correlations between the probability of C-section (ρ_i^r) and doctor characteristics in Table 3. Table 3 shows that the correlation between a high risk of C-section and our measures of diagnostic skill and procedural skill are quite low (.055 and -.033, respectively). However, there is a correlation of .160 between ρ_i^r and the share high risk in the practice, suggesting once again that high risk women tend to 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 .259. And there are sizable negative correlations between the rate of bad outcomes in a practice and our two skill measures, which is reassuring: The correlation between the rate of bad outcomes and our measure of diagnostic skill is -.283, while the correlation between bad outcomes and our measure of procedural skill is -.446. This analysis suggests that controlling for the share high risk in the practice is one way to control for an important observable aspect of selection.

The other ways that we address the selection issue are as follows: (1) 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 (2) 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) our definition of a hospital market is defined with referenct to 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 residence, plus any hospital used by more than three women from her zip code of residence in the birth year.[8] Thus, there is a distinct market, or set of hospital choices, facing each woman at the time of each birth.

Figure 4 shows the distribution of hospitals and illustrates this way of defining markets. The figure shows that most women choose nearby hospitals, 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, in the appendix we also estimate models using only first births. The idea behind these models is that mothers (and doctors) have much less information about likely outcomes for a first birth than for subsequent births and so may be less selective about physicians.

5 Results

Table 4 shows estimates of equation (20), where the dependent variable is whether there was a C-section. Table 4 indicates that diagnostic skill and procedural skill have distinct effects. When 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 the two lowest risk groups and increases it for the two other groups with an especially large effect in the highest risk group. A larger price gap between

[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.

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. Percentage changes calculated in this way are shown in Appendix Table 3. For the index of diagnostic skill, this movement (of .215 units) would reduce C-section rates by 15.8% among the lowest risk, and by 10.7% among the low risk, but would increase them by 3.8% and 4.7% among medium and high risk women respectively. These figures imply a large overall decrease in C-section rates with better diagnosis. Specifically, they imply a net decrease of 35,507 women receiving C-sections, which is about 3.7% of the births in our sample.

For the index of procedural skill, a movement from the 25th to the 75th percentile of the distribution (a movement of .062 units) would increase the probability of C-section by 3.7%, 3.8%, .8% and .5% for very low, low, medium, and high risk women, respectively. Finally, the estimates for prices 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 5.4% in the very low risk group, 8.8% among the low risk, and 3.2% among the medium risk, but would have no impact on the high risk, where medical necessity is a much more important determinant of C-sections 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 high C-section rates are a response to demand from white, college-educated women. Instead, conditional on medical risk, 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 17.0%, 9.9%, 10.6%, and 8.3% among the very low, low, medium, and high risk, respectively. The incidence of neonatal death also declines significantly (though since neonatal deaths are a rare outcome, the implied 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 6.5%, 17.7%, 20.4%, and 55.7% 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 also large and increasing in medical risk. An increase in the price gap of $2,600 has no statistically significant effect on the probability of any bad outcome (though the point estimate is positive), but is estimated to increase the risk of neonatal death among all but the highest risk group. A price increase is estimated to have no effect on the risk of death among the high risk, which 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. Fortunately, our analysis of observable characteristics of doctors and patients suggests that any selection that is occuring is of the first type, and therefore that our estimates are likely to understate the effects of doctor 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. 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 panel of Table 7 to Table 4 indicates that the estimated effect of diagnosis on the probability of C-section remains statistically significant though the magnitude of the estimates are affected by the exclusion of planned C-sections. Some of the planned C-sections may be the cases where diagnostic criterion that we do not observe dictate a C-section. In all but the highest risk group the effect of diagnostic skill is reduced by the exclusion of planned C-sections. 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 panels 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 very low risk patients.[10] In order to compute this measure, we take the mean C-section rate for high risk patients in the market, and the mean C-section rate for very low risk patients in the market and subtract. This measure has a mean of 0.830 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 analogously to the way it

[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.

was defined 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 19.1% decline in C-sections among the very low risk and with an increase of 4.2% in the probability of C-sections among the high risk. At the same time, better diagnosis is estimated to significantly decrease the probability of bad outcomes among all risk groups. The point estimates on the measure of diagnostic skill suggest that an improvement of this magnitude would also lower neonatal deaths, especially among the very low risk and the very high risk groups, but these coefficients are not precisely estimated.

An improvement in surgical skill relative to skill at natural delivery that moved a physician from the 25th to the 75th percentile of the distribution is estimated to increase the probability of C-section (by 5.4% for the very low risk, 1.4% for the next two risk groups, and by 0.8% for the highest risk group). It would also reduce the incidence of bad outcomes for all but the very low risk group, where it would increase the risk of bad outcomes (presumably by encouraging unnecessary C-sections and risking surgical complications). An increase in the price gap between C-sections and natural deliveries is estimated to have the greatest effect on C-sections among those in the two lower risk groups, increasing the incidence of C-section by 5% and 7%.

Appendix Table 1 shows the results of estimating our models only on the sample of first births. The idea is that doctors and patients may have less basis for selecting a physician at a first birth than for subsequent births. As Table 2 showed, very few first births are in the lowest risk category, hence we do not estimate the model separately for this risk category. Appendix Table 1 shows that the results are qualitatively similar to those in Tables 4 and 5. The effect of procedural skills and prices on the probability of C-section are somewhat higher than in the full sample. For outcomes, the estimates are quite similar to those in Tables 4 and 5, while prices seem to have a smaller impact on first births than in the full sample.

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).

6 Discussion and Conclusions

The previous literature on treatment choice emphasizes that it is affected by physician skill, but only allows physician skill to vary along a single dimension which can be thought of as technical skill in executing procedures. Taking a cue from the literature on managerial decision making (Bloom and Van Reenen (2010)) we develop a model that includes an additional dimension of skill: Diagnostic decision making. In our model, a good doctor is one who is not only technically skilled, but is also able to draw the correct inferences from the available data in order to match patients correctly to the procedures that are most likely to benefit them. This simple framework yields rich predictions and allows us to distinguish between the two types of skill. The model shows that better procedural skill leads to higher use of intensive procedures across the board, for both high and low risk patients. This finding yields the possibility, confirmed in our data, that improvements in procedural skill could actually harm the lowest risk patients by making it more likely that they will be subjected to unnecessarily intensive procedures that can only harm them. In contrast, 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 thus leads to better health outcomes in both groups.

We provide an application of our model using 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 15.8% among the very low risk, and increase them by 4.7% 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 benefits for high risk women.

Our work highlights the importance of diagnostic decision making skill in medicine and suggests an empirical approach to measuring it. As such, it constitutes a first step towards improving diagnostic decision making skill. Future research into the mechanisms that could best accomplish this goal is warranted.

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.

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics 116*(1), 313–50.

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.

Bloom, N. and J. Van Reenen (2010). Human resource managment and productivity. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Volume 4*, Volume 4, Chapter 19. 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.

Epstein, 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.

Farber, H. S. and R. Gibbons (1996, November). Learning and wage dynamics. *Quarterly Journal of Economics 111*(4), 1007–47.

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.

Grant, D. (2009). Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project. *Journal of Health Economics 28*, 244–250.

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.

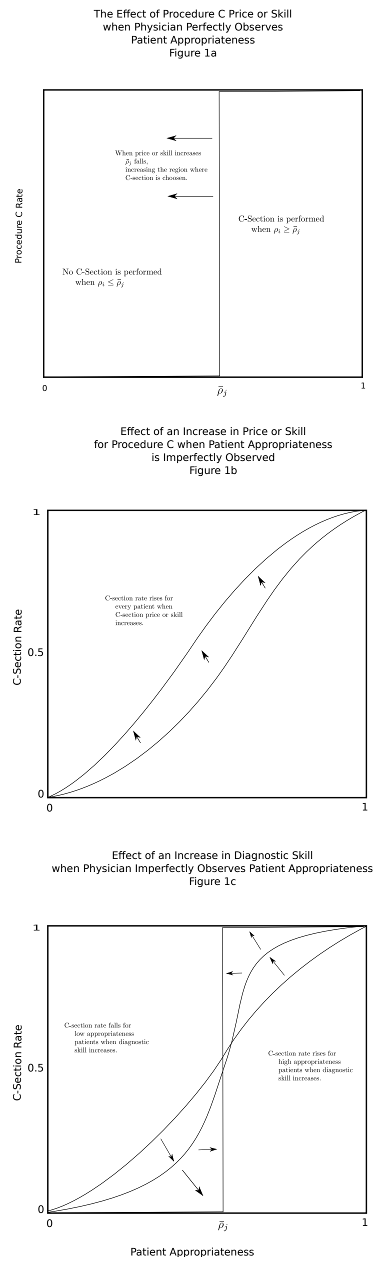


Figure 1: Effect of Patient Appropriateness upon Physician Choice

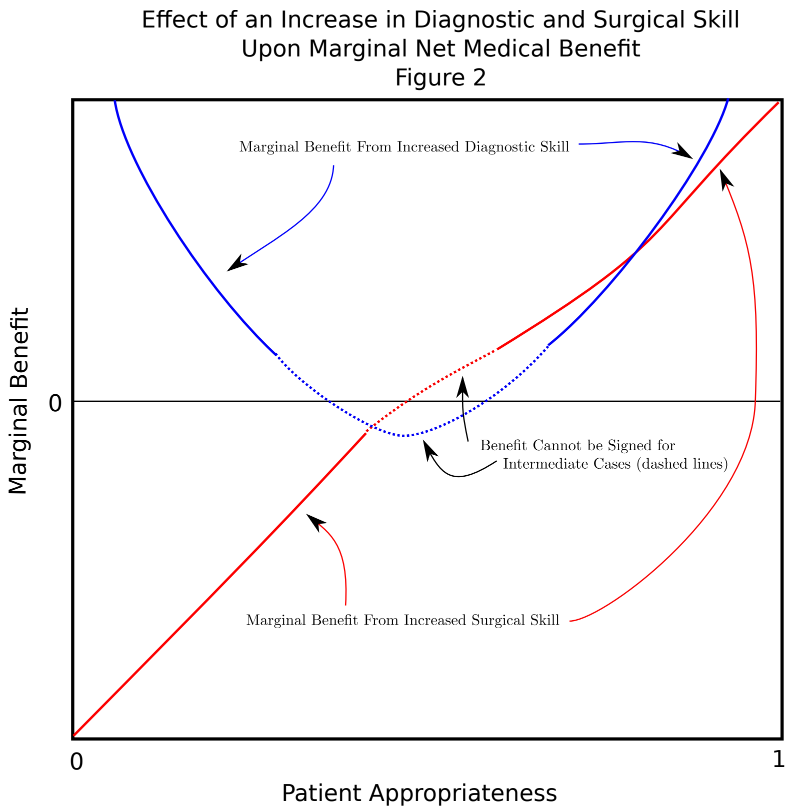
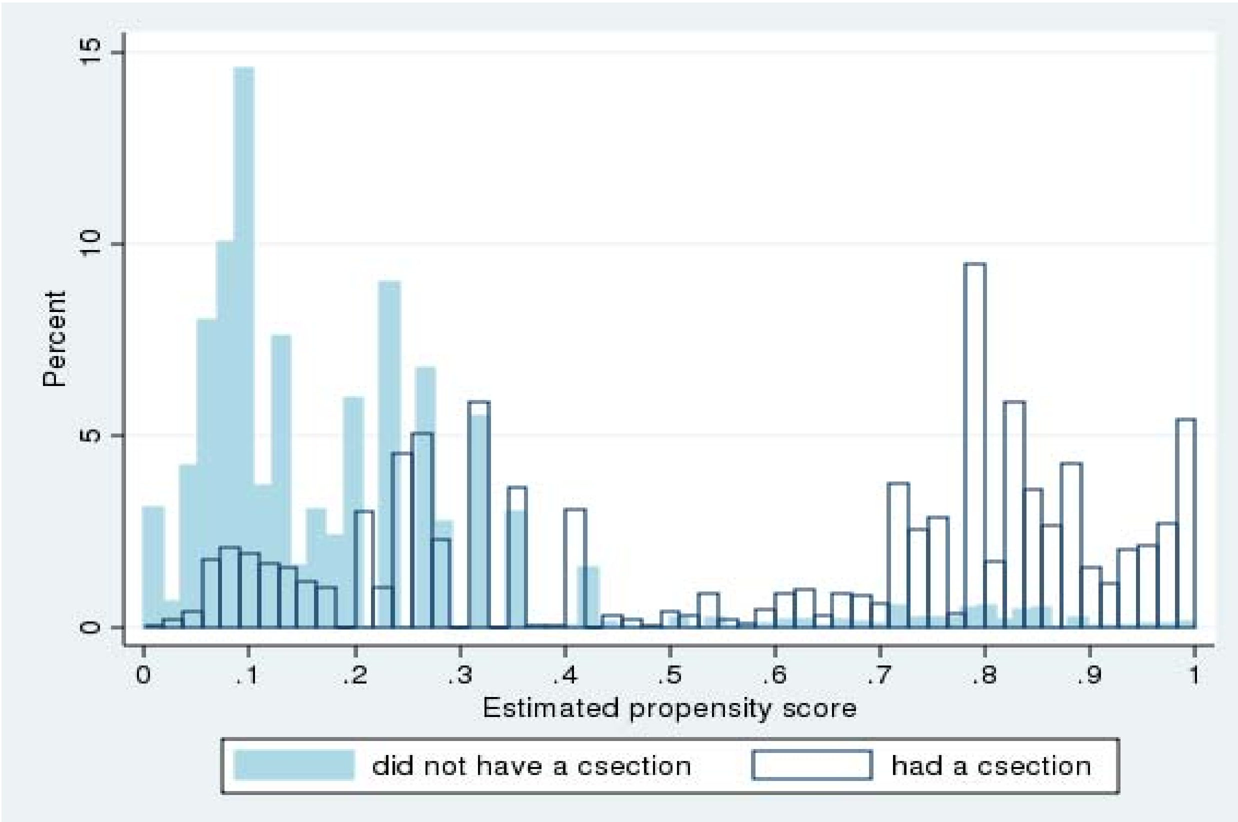


Figure 2: Effect of Skill on Net Medical Benefit

Figure 3: Predicting C-sections Using the Logit Model



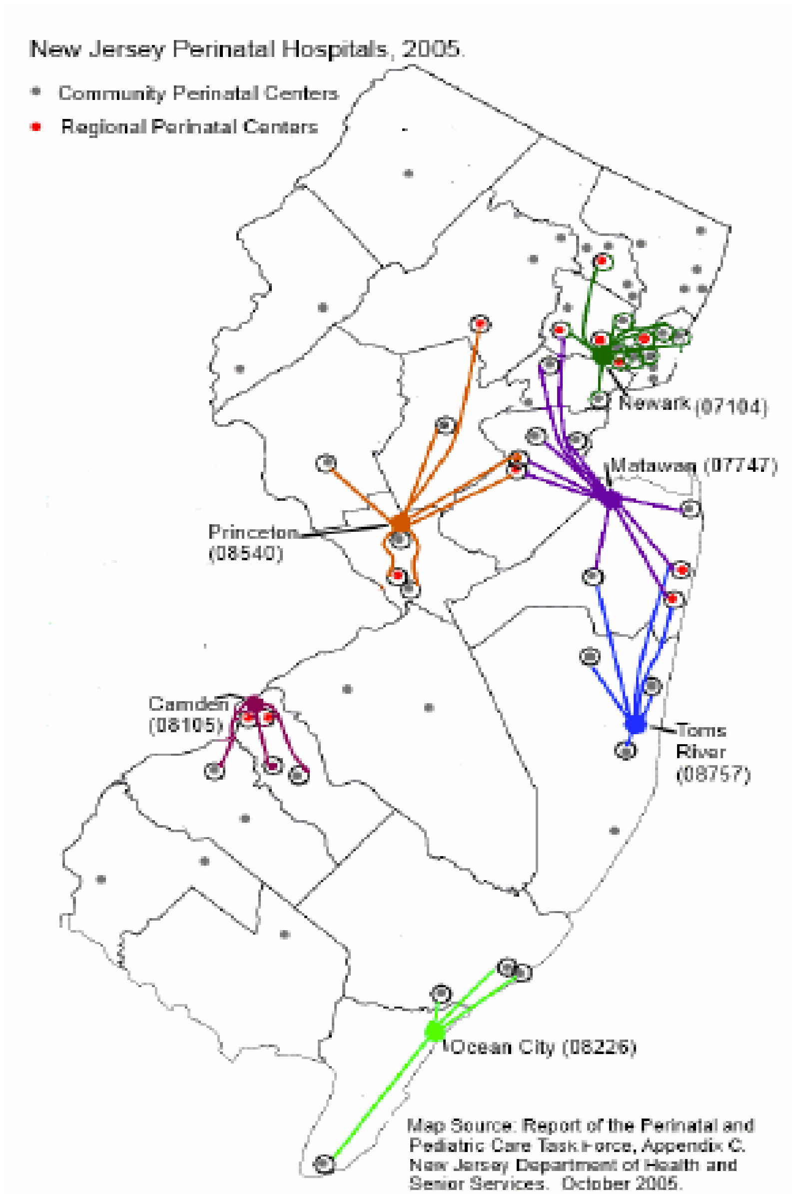


Figure 4: Illustrating the Definition of a Market

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

	All Doctors			Good Doctors Only		
	Coeff.	S.E.	Marginal Effect	Coeff.	S.E.	Marginal Effect
Age<20	-0.337	0.013	-0.075	-0.428	0.029	-0.095
Age >=25&<30	0.262	0.008	0.058	0.311	0.018	0.069
Age >=30&<35	0.434	0.008	0.096	0.483	0.017	0.107
Age >=35	0.739	0.009	0.164	0.840	0.018	0.186
2nd Birth	-1.347	0.007	-0.298	-1.448	0.015	-0.321
3rd Birth	-1.645	0.009	-0.364	-1.787	0.019	-0.396
4th or Higher Birth	-2.140	0.012	-0.474	-2.317	0.027	-0.513
Previous C-section	3.660	0.008	0.810	3.885	0.018	0.860
Previous Large Infant	0.139	0.029	0.031	0.293	0.065	0.065
Previous Preterm	-0.293	0.025	-0.065	-0.311	0.061	-0.069
Multiple Birth	2.879	0.014	0.638	3.278	0.032	0.726
Breech	3.353	0.016	0.742	3.810	0.040	0.844
Placenta Previa	3.811	0.054	0.844	3.843	0.116	0.851
Abruptio Placenta	2.048	0.030	0.454	2.196	0.072	0.486
Cord Prolapse	1.761	0.047	0.390	1.668	0.100	0.369
Uterine Bleeding	0.026	0.035	0.006	0.259	0.099	0.057
Eclampsia	1.486	0.096	0.329	1.047	0.230	0.232
Chronic Hypertension	0.745	0.025	0.165	0.754	0.060	0.167
Pregnancy Hypertension	0.639	0.013	0.142	0.696	0.029	0.154
Chronic Lung Condition	0.064	0.014	0.014	0.110	0.032	0.024
Cardiac Condition	-0.121	0.020	-0.027	-0.175	0.042	-0.039
Diabetes	0.558	0.011	0.124	0.547	0.025	0.121
Anemia	0.131	0.018	0.029	0.203	0.043	0.045
Hemoglobinopathy	0.116	0.047	0.026	0.067	0.092	0.015
Herpes	0.461	0.024	0.102	0.558	0.049	0.124
Other STD	0.052	0.017	0.012	0.064	0.039	0.014
Hydramnios	0.616	0.018	0.136	0.645	0.042	0.143
Incompetent Cervix	0.043	0.035	0.010	-0.119	0.093	-0.026
Renal Disease	-0.024	0.031	-0.005	-0.057	0.067	-0.013
Rh Sensitivity	-0.045	0.040	-0.010	-0.082	0.109	-0.018
Other Risk Factor	0.276	0.006	0.061	0.210	0.013	0.047
Constant	-1.414	0.007	-0.313	-1.374	0.015	-0.304
# Observations	1169654			262174		
Pseudo R2	0.32			0.322		

Notes: The model also included indicators for missing age, parity, and risk factors.
The correlation between rho estimated using the two different models is .99.

Table 2: Means by Probability of C-Section

Medical Risk:	All	Very Low	Low	Medium	High
<u>Outcomes</u>					
C-Section Rate	0.331	0.06	0.115	0.439	0.891
Any Bad Outcome	0.066	0.038	0.048	0.085	0.086
Neonatal death (per 1000)	4.082	3.222	2.380	4.778	7.850
<u>Doctor Characteristics</u>					
# Deliveries in Sample	1019.45	1048.14	1025.22	1008.63	1011.13
	(650.15)	(737.95)	(655.31)	(625.62)	(627.24)
Diagnostic Skill	1.033	1.005	1.034	1.039	1.037
	(0.089)	(0.191)	(0.183)	(0.180)	(0.181)
Procedural Skill Differential	-0.049	-0.055	-0.049	-0.047	-0.050
	(0.065)	(0.071)	(0.066)	(0.064)	(0.064)
Price Differential ($1000)	4.755	5.066	4.748	4.694	4.702
	(2.678)	(2.850)	(2.675)	(2.635)	(2.643)
Share High Risk	0.122	0.107	0.118	0.124	0.137
	(0.043)	(0.040)	(0.041)	(0.043)	(0.048)
Rate of Bad Outcomes	0.066	0.067	0.065	0.066	0.069
	(0.042)	(0.045)	(0.041)	(0.041)	(0.045)
<u>Mother & Child Characteristics</u>					
African American	0.158	0.257	0.164	0.132	0.134
Hispanic	0.210	0.315	0.224	0.179	0.180
Married	0.713	0.533	0.678	0.770	0.796
High School Dropout	0.128	0.289	0.145	0.082	0.082
Teen mom	0.030	0.026	0.060	0.008	0.013
Mom Age 35 or More	0.238	0.139	0.244	0.197	0.439
Smoked	0.081	0.129	0.079	0.013	0.074
Child Male	0.513	0.514	0.514	0.513	0.514
Child First Born	0.398	0.016	0.252	0.676	0.284
Medicaid	0.206	0.389	0.223	0.156	0.152
# of Observations	968748	104902	364268	381745	117833

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

	P(C-section)	# Deliveries	Diagnostic Skill	Procedural Skill	Price Difference	Share High Risk
P(C-section)	1					
# Deliveries	-0.009	1				
Diagnostic Skill	0.016	0.044	1			
Procedural Skill Diff.	-0.003	0.053	0.259	1		
Price Difference	-0.005	0.048	-0.017	-0.032	1	
Share High Risk	0.161	-0.036	0.055	-0.033	-0.119	1
Rate of Bad Outcomes	0.034	-0.096	-0.283	-0.446	0.013	0.199

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

Medical Risk:	Very Low	Low	Medium	High
Diagnostic Skill	-0.044	-0.057	0.077	0.194
	(0.008)	(0.009)	(0.013)	(0.010)
Procedural Skill Difference	0.036	0.071	0.055	0.072
	(0.017)	(0.021)	(0.037)	(0.029)
Price Differential (x 100)	0.125	0.389	0.537	-0.018
	(0.051)	(0.063)	(0.095)	(0.054)
C-section Risk	0.190	1.062	0.870	0.794
	(0.090)	(0.037)	(0.010)	(0.021)
African-American	0.023	0.057	0.065	0.025
	(0.004)	(0.003)	(0.006)	(0.005)
Hispanic	0.009	0.027	0.057	0.037
	(0.004)	(0.003)	(0.005)	(0.004)
Less than High School	0.013	0.023	0.030	0.015
	(0.004)	(0.003)	(0.004)	(0.005)
High School	0.019	0.026	0.039	0.021
	(0.004)	(0.002)	(0.003)	(0.003)
Some College	0.010	0.012	0.018	0.006
	(0.004)	(0.002)	(0.002)	(0.003)
Married	-0.009	-0.011	-0.009	0.005
	(0.002)	(0.002)	(0.002)	(0.003)
Medicaid	0.002	0.009	0.0000	0.012
	(0.002)	(0.003)	(0.004)	(0.003)
Teen Mom	-0.017	-0.024	0.007	0.015
	(0.005)	(0.004)	(0.010)	(0.010)
Mother 25-34	0.022	0.038	0.015	0.005
	(0.002)	(0.004)	(0.004)	(0.004)
Mother 35+	0.044	0.042	0.035	0.006
	(0.004)	(0.005)	(0.004)	(0.004)
Mother Smoked	0.007	0.012	0.005	-0.003
	(0.002)	(0.002)	(0.003)	(0.004)
Child Male	0.010	0.021	0.033	0.005
	(0.001)	(0.001)	(0.002)	(0.002)
Child 2nd Born	-0.111	-0.032	0.091	0.001
	(0.009)	(0.006)	(0.005)	(0.004)
Child 3rd Born	-0.126	-0.029	0.063	-0.032
	(0.009)	(0.006)	(0.005)	(0.004)
Child 4th Born or Higher	-0.124	0.002	0.025	-0.050
	(0.010)	(0.007)	(0.007)	(0.004)
R-squared	0.015	0.041	0.219	0.059
# Observations	104902	364268	381745	117833

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:	**Very Low**	**Low**	**Medium**	**High**	**Very Low**	**Low**	**Medium**	**High**
Diagnostic Skill	-0.030	-0.022	-0.042	-0.033	-0.005	-0.003	-0.010	-0.017
	(0.007)	(0.005)	(0.008)	(0.011)	(0.001)	(0.001)	(0.001)	(0.002)
Procedural Skill Difference	0.040	-0.137	-0.280	-0.772	0.007	-0.008	-0.016	-0.057
	(0.025)	(0.021)	(0.031)	(0.034)	(0.004)	(0.002)	(0.004)	(0.008)
Price Differential (x 100)	0.082	0.040	-0.003	0.029	0.013	0.008	0.025	-0.014
	(0.053)	(0.041)	(0.052)	(0.054)	(0.007)	(0.003)	(0.007)	(0.012)
C-section Risk	0.076	0.293	-0.038	0.378	-0.108	0.014	-0.013	0.0470
	(0.068)	(0.022)	(0.006)	(0.019)	(0.026)	(0.006)	(0.002)	(0.007)
R-squared	0.006	0.013	0.010	0.053	0.004	0.004	0.008	0.016
# Observations	104902	364268	381745	117833	104902	364268	381745	117833

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

Medical Risk:	Very Low	Low	Medium	High
C-Section				
Diagnostic Skill	-0.05	-0.063	0.076	0.197
	(0.007)	(0.009)	(0.013)	(0.010)
Procedural Skill Difference	0.045	0.083	0.079	0.092
	(0.016)	(0.021)	(0.036)	(0.029)
Price Differential (x 100)	0.154	0.42	0.599	0.013
	(0.050)	(0.062)	(0.092)	(0.053)
C-section Risk	0.156	1.034	0.864	0.772
	(0.090)	(0.036)	(0.010)	(0.021)
Share High Risk in Practice	0.299	0.332	0.561	0.264
	(0.035)	(0.044)	(0.077)	(0.037)
R-squared	0.0178	0.043	0.221	0.061
Any Bad Outcome				
Diagnostic Skill	-0.034	-0.024	-0.042	-0.031
	(0.006)	(0.005)	(0.007)	(0.010)
Procedural Skill Difference	0.044	-0.132	-0.272	-0.760
	(0.024)	(0.020)	(0.029)	(0.033)
Price Differential (x 100)	0.096	0.054	0.017	0.047
	(0.052)	(0.040)	(0.051)	(0.053)
C-section Risk	0.060	0.280	-0.040	0.365
	(0.068)	(0.022)	(0.006)	(0.018)
Share High Risk in Practice	0.145	0.151	0.194	0.149
	(0.030)	(0.023)	(0.033)	(0.040)
R-squared	0.007	0.014	0.011	0.053
Neonatal Death				
Diagnostic Skill	-0.006	-0.003	-0.010	-0.017
	(0.001)	(0.001)	(0.001)	(0.002)
Procedural Skill Difference	0.007	-0.008	-0.015	-0.056
	(0.004)	(0.002)	(0.004)	(0.008)
Price Differential (x 100)	0.014	0.009	0.028	-0.012
	(0.007)	(0.004)	(0.006)	(0.012)
C-section Risk	-0.108	0.013	-0.013	0.045
	(0.026)	(0.006)	(0.002)	(0.007)
Share High Risk in Practice	0.005	0.010	0.026	0.017
	(0.007)	(0.004)	(0.007)	(0.010)
R-squared	0.004	0.004	0.009	0.016
# Observations	104902	364268	381745	117833

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

Medical Risk:	Very Low	Low	Medium	High
C-Section				
Diagnostic Skill	-0.022	-0.033	0.019	0.342
	(0.005)	(0.007)	(0.013)	(0.020)
Procedural Skill Difference	0.020	0.057	0.057	0.098
	(0.013)	(0.016)	(0.029)	(0.052)
Price Differential (x 100)	0.062	0.232	0.373	-0.219
	(0.037)	(0.048)	(0.095)	(0.130)
C-section Risk	0.211	0.530	0.508	2.269
	(0.066)	(0.029)	(0.010)	(0.049)
R-squared	0.012	0.046	0.037	0.192
Any Bad Outcome				
Diagnostic Skill	-0.027	-0.020	-0.040	-0.051
	(0.006)	(0.005)	(0.008)	(0.015)
Procedural Skill Difference	0.037	-0.129	-0.229	-0.831
	(0.025)	(0.021)	(0.031)	(0.044)
Price Differential (x 100)	0.008	0.004	-0.036	-0.077
	(0.053)	(0.041)	(0.057)	(0.102)
C-section Risk	0.057	0.268	0.020	0.369
	(0.067)	(0.023)	(0.007)	(0.034)
R-squared	0.006	0.013	0.006	0.045
Neonatal Death				
Diagnostic Skill	-0.004	-0.002	-0.009	-0.026
	(0.001)	(0.001)	(0.001)	(0.004)
Procedural Skill Difference	0.007	-0.008	-0.012	-0.083
	(0.004)	(0.002)	(0.004)	(0.015)
Price Differential (x 100)	0.013	0.007	0.035	-0.014
	(0.007)	(0.003)	(0.008)	(0.030)
C-section Risk	-0.110	0.003	-0.006	0.071
	(0.025)	(0.005)	(0.002)	(0.016)
R-squared	0.004	0.004	0.012	0.025
# Observations	104902	364268	381745	117833

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

Medical Risk:	Very Low	Low	Medium	High
C-Section				
Diagnostic Skill	-0.166	-0.157	0.062	0.542
	(0.051)	(0.034)	(0.050)	(0.068)
Procedural Skill Difference	0.111	0.054	0.207	0.248
	(0.055)	(0.049)	(0.065)	(0.075)
Price Differential (x 100)	0.116	0.311	0.300	-0.158
	(0.049)	(0.056)	(0.066)	(0.050)
C-section Risk	0.175	1.074	0.870	0.783
	(0.085)	(0.035)	(0.010)	(0.021)
R-squared	(0.024)	(0.048)	(0.224)	(0.057)
Any Bad Outcome				
Diagnostic Skill	-0.103	-0.116	-0.102	-0.122
	(0.038)	(0.025)	(0.029)	(0.048)
Procedural Skill Difference	0.080	-0.124	-0.181	-0.543
	(0.047)	(0.036)	(0.043)	(0.068)
Price Differential (x 100)	0.093	0.060	0.132	0.163
	(0.054)	(0.045)	(0.051)	(0.065)
C-section Risk	0.024	0.300	-0.040	0.402
	(0.065)	(0.019)	(0.006)	(0.018)
R-squared	0.020	0.019	0.013	0.037
Neonatal Death				
Diagnostic Skill	-0.006	0.000	-0.007	-0.014
	(0.013)	(0.004)	(0.007)	(0.014)
Procedural Skill Difference	0.019	0.006	0.003	-0.047
	(0.013)	(0.006)	(0.011)	(0.021)
Price Differential (x 100)	0.011	0.009	0.032	-0.006
	(0.012)	(0.005)	(0.007)	(0.019)
C-section Risk	-0.118	0.015	-0.013	0.051
	(0.026)	(0.005)	(0.002)	(0.006)
R-squared	0.017	0.005	0.010	0.018
# Observations	104902	364268	381745	117833

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

Online Appendix - Not For Publication

1 Appendix

In the text we provide a new, empirical measure of diagnostic skill defined by the extent to which a physician's treatment decision varies with observed patient condition, h_i^r. If the variations in diagnostic skill are due to variations in the ability to use the information in h_i^r then an increase in sensitivity to h_i^r results in better outcomes for the patient. We find that this interpretation is consistent with the evidence.

This is not the only interpretation possible. In fact, a standard economic model would typically assume that since h_i^r is defined using information from the medical record which is available to physicians, then physicians use this information efficiently. In that case, physicians with better diagnostic skill observe h_i^r plus some other information that we cannot see. In this appendix we show that this model has the *opposite* prediction to what we find in the data.

To see this suppose that h_i^r is very low, and hence from our data the patient is a poor candidate for a C-section. During labor the physician may observe conditions that are *not* reported in our data, but indicate that a C-section would be appropriate. This would imply that the physician should ignore h_i^r and put more weight on this private information. Empirically this would imply a decrease in the sensitivity of the doctor's decision about procedure choice with respect to h_i^r. Hence, in the case considered in this appendix, a physician with better diagnostic skill should be less sensitive to h_i^r to achieve better outcomes. The next subsection derives this implication formally.

1.1 Effect of Diagnostic Skill When h_i^r is Perfectly Observed

As before, suppose that the physician's *ex ante* distribution for the true patient (log) condition, h_i, has distribution:

$$h_i \sim N\left(h_j^0, 1/\rho_j^0\right),$$

where $\rho_j^0 = (D_j^0)^2$ is the precision of these beliefs, and D_j^0 is what we have denoted above as the strength of physician beliefs, or the "dogmatism" of the physician. From our data we have a signal regarding the condition of the patient given by:

$$h_i^r = h_i + \epsilon_i^r,$$

where $\epsilon_i^r \sim N\left(0, \sigma_r^2\right)$. Here h_r^i is an unbiased signal of patient condition (C-section appropriateness) that we have estimated using the data from all of New Jersey which has precision $D_r^2 = \rho^r = 1/\sigma_r^2$.

In the text (equation 7) we suppose that the physician observes:

$$h_{ij} = h_i^r + \epsilon_{ij}/D_j. \tag{1}$$

In addition, let us suppose that this doctor has some additional information we do not see:

$$\bar{h}_{ij} = h_i + \bar{\epsilon}_{ij}/\bar{D}_j,$$

where $\bar{\epsilon}_{ij} \sim N(0,1)$, and the precision of this estimate is $\bar{\rho}_j = \bar{D}_j^2$. In this case the doctor has information set:

$$I_{ij} = \left\{h_{ij}, \bar{h}_{ij}\right\}.$$

The model in the text is a special case of this model where$\bar{D}_j = 0$, that is, the physician has no additional information. In that case our measure of diagnostic skill is identical to D_j.

In order to contrast the effect of this new information, let us suppose that the physician observes h_i^r perfectly - this corresponds to $D_j \to \infty$, while $\bar{D}_j \in (0, \infty)$ is allowed to vary between physicians. We will now derive the effect of $\bar{D}_j$ upon our measure of diagnostic skill, and show that our proposed measure is still a clean measure of information processing skill that is independent of procedural skill. Second, we show that in this case an improvement in information processing leads to a fall in the diagnostic skill measure.

With this information we can now apply the rule for Bayesian learning to compute the physicians' beliefs regarding patient condition. From DeGroot (1972) we have the familiar learning rule:

$$E\left\{h_i|\bar{h}_{ij}, h_j^0, h_i^r\right\} = \frac{\bar{\rho}_j}{\rho_j^*}\bar{h}_{ij} + \frac{\rho_j^0}{\rho_j^*}h_j^0 + \frac{\rho^r}{\rho_j^*}h_i^r \tag{2}$$

where $\rho_j^* = \bar{\rho}_j + \rho_j^0 + \rho^r$ is the *posterior* precision of the physician's estimate of patient condition. Notice that this posterior precision - effectively how sure the doctor is regarding the patient's status - varies with both prior beliefs and the quality of her personal information as measured by $\bar{D}_j$.

From expression 10 procedure C is chosen if and only if:

$$E\left\{h_i|\bar{h}_{ij}, h_j^0, h_i^r\right\} + s_j^C - s_j^N + m_j\left(\Delta P\right) \geq 0.$$

Hence we observe procedure C if and only if:

$$\frac{\bar{\rho}_j}{\rho_j^*}\bar{h}_{ij} + \frac{\rho_j^0}{\rho_j^*}h_j^0 + \frac{\rho^r}{\rho_j^*}h_i^r + s_j^C - s_j^N + m_j\left(\Delta P\right) \quad \geq 0. \tag{3}$$

Since we have panel data on physician decision making we suppose that we can estimate physician specific parameters. However we cannot observe $\bar{h}_{ij}$. Rather, observe:

$$\bar{h}_{ij} = h_i + \epsilon_{ij}/\bar{D}_j \tag{4}$$

$$= h_i^r - \epsilon_i^r + \epsilon_{ij}/\bar{D}_j. \tag{5}$$

Thus, the decision rule can be rewritten as:

$$\frac{\bar{\rho}_j}{\rho_j^*}h_i^r + \frac{\rho_j^0}{\rho_j^*}h_j^0 + \frac{\rho^r}{\rho_j^*}h_i^r + s_j^C - s_j^N + m_j\left(\Delta P\right) \geq \frac{\bar{\rho}_j}{\rho_j^*}\left(\epsilon_i^r/D_r - \epsilon_{ij}/\bar{D}_j\right).$$

This expression can be rewritten as:

$$v_j\left(\bar{D}_j\right)\left(h_i^r + b_j\left(\bar{D}_j\right)\right) \geq \epsilon_{ij}^2, \tag{6}$$

where ϵ_{ij}^2 has a standard normal distribution that is independent across observations. (Compare this with expression 11 above). The function $v_j\left(\bar{D}_j\right)$ corresponds to our measure of diagnostic skill, while the intercept term $b_j\left(\bar{D}_j\right)$ is the cutoff point at which a patient is more likely to get a C-section.

Recalling that $\bar{D}_j^2 = \bar{\rho}_j$ the effect of $\bar{D}_j$ upon our measure of diagnostic skill is given by:

$$v_j\left(\bar{D}_j\right) = \frac{1 + \rho^r/\bar{\rho}_j}{1 + \bar{\rho}_j},$$

$$= \frac{1 + \rho^r/\bar{D}_j^2}{1 + \bar{D}_j^2}.$$

Notice that this implies that $dv_j(D_j)/dD_j < 0$, and hence when the physician has private information her decisions about procedure choice are less sensitive to h_i^r. The intercept term is given by:

$$
\begin{aligned}
b_j\left(\bar{D}_j\right) &= \frac{\rho_j^0 h_j^0 + \rho_j^* \left(s_j^C - s_j^N + m_j\left(\Delta P\right)\right)}{\bar{\rho}_j + \rho^r}, && (7)\\
&= \frac{\rho_j^0 h_j^0 + \left(D_j^2 + \rho_j^0 + \rho^r\right)\left(s_j^C - s_j^N + m_j\left(\Delta P\right)\right)}{\bar{D}_j^2 + \rho^r}, && (8)\\
&= \frac{\rho_j^0}{\bar{D}_j^2 + \rho^r}\left(h_j^0 + s_j^C - s_j^N + m_j\left(\Delta P\right)\right) && (9)\\
&\quad + \left(s_j^C - s_j^N + m_j\left(\Delta P\right)\right). && (10)
\end{aligned}
$$

It is readily verified that we cannot sign the effect of $\bar{D}_j$ upon the intercept term. Thus we have:

Proposition 1. *Holding all else fixed, the observed probability that the physician uses procedure C decreases with the precision with which they observe true patient condition (h_i) if and only if observed patient condition (h_i^r) is above a fixed, physician specific, threshold ($h_i^r > -b_j(D_j)$).*

This result shows that when the physician has better information than the information reported on the birth record, then our measure of diagnostic skill falls as this information is improved.

The next issue is the effect of this information upon patient outcomes. We work this out by considering the problem from the perspective of a patient who has condition h_i and is attended to by physician j. The issue is whether or not an increase in the precision of this physician's information (D_j), will improve the expected medical outcome.

The expected medical outcome can be written as a function of patient condition and the likelihood of procedure C or N:

$$
\begin{aligned}
NB_j(h_i) = &\quad \left(h_i + s_j^C - s_j^N\right) Pr_j\left[C|h_i\right] \\
- &\quad \left(h_i + s_j^C - s_j^N\right) Pr_j\left[N|h_i\right] \\
= &\quad \left(h_i + s_j^C - s_j^N\right)\left(2Pr_j\left[C|h_i\right] - 1\right).
\end{aligned}
$$

Notice that if the health condition of the patient were observed perfectly, then procedure C is performed if and only if $\left(h_i + s_j^C - s_j^N\right) \geq 0$.

Using expression 6 we can compute the probability of procedure C as a function of h_i. Substituting in the random terms we get that procedure C carried out if and only if:

$$
\bar{\rho}_j(h_i + \epsilon_{ij}/D_j) + \rho_j^0 h_j^0 + \rho^r(h_i + \epsilon_i^r/D_r) + \rho_j^*\left(s_j^C - s_j^N + m_j\left(\Delta P\right)\right) \geq 0.
$$

This is rewritten as:

$$
\frac{(\bar{\rho}_j + \rho^r)h_i + \rho_j^0 h_j^0}{\rho_j^*} + \left(s_j^C - s_j^N + m_j\left(\Delta P\right)\right) \geq \frac{(D_j\epsilon_{ij} + D_r\epsilon_i^r)}{\rho_j^*}.
$$

The term on the right is an i.i.d. normally distributed random variable with mean zero and variance $\left(\frac{\bar{\rho}_j+\rho^r}{\left(\rho_j^*\right)^2}\right)$. From this we conclude that:

$$Pr_j\left[C|h_i\right] = F\left(\rho_j^*\left(h_i + \frac{\rho_j^0}{\bar{\rho}_j+\rho^r}h_j^0 + s_j^C - s_j^N + m_j\left(\Delta P\right)\right)\right). \tag{11}$$

Notice that as diagnostic information increases ($\bar{\rho}_j \to \infty$), then the decision rule approaches one in which procedure C is chosen if and only if:

$$h_i + s_j^C - s_j^N + m_j\left(\Delta P\right) \geq 0,$$

which is the optimal choice when the cost of care $m_j\left(\Delta P\right)$ is taken into account.

An increase in precision $\bar{D}_j$ improves the medical outcome if and only if:

$$\left(h_i + s_j^C - s_j^N\right)\frac{\partial Pr_j\left[C|h_i\right]}{\partial \bar{D}_j} \geq 0.$$

This is positive if and only if:

$$\left(h_i + s_j^C - s_j^N\right)F'\left(h_i + s_j^C - s_j^N + \left(m_j\left(\Delta P\right) + \frac{\rho_j^0}{\bar{\rho}_j+\rho^r}h_j^0\left(1 - \frac{1}{\bar{\rho}_j+\rho^r}\right)\right)\right) \geq 0. \tag{12}$$

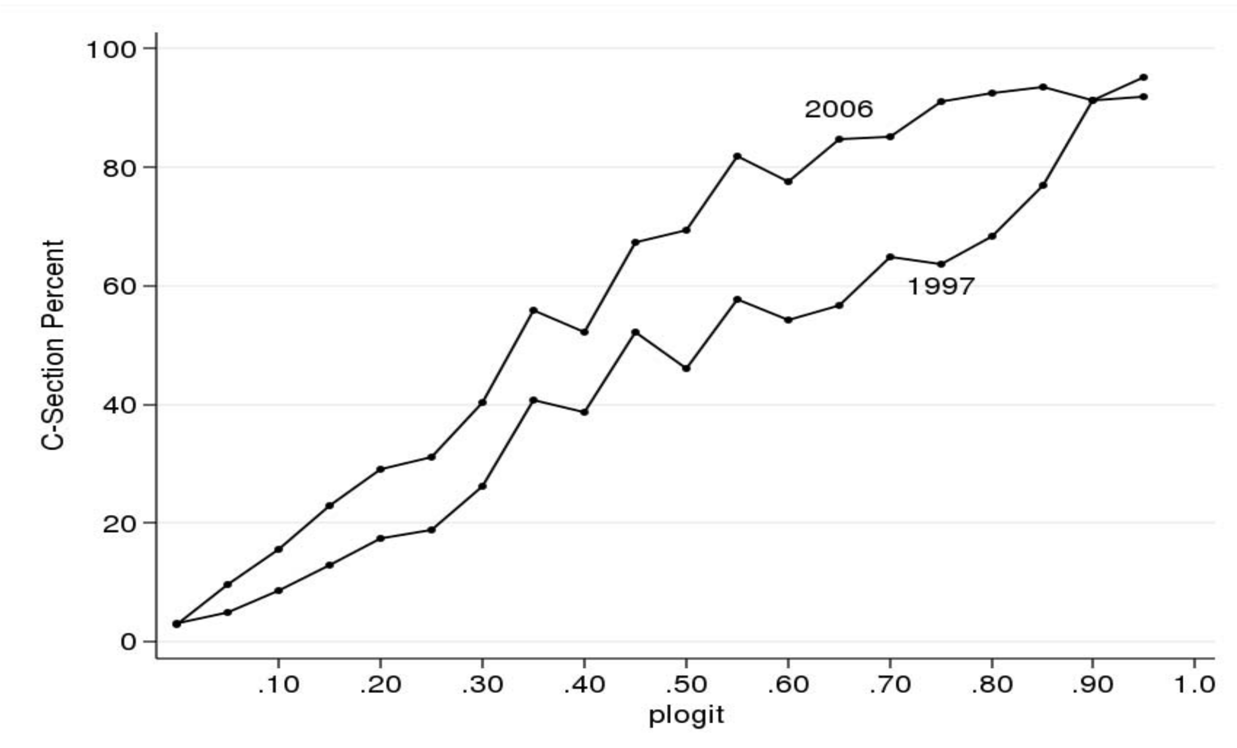
From this we get the following proposition.

Proposition 2. *When the patient condition is either very appropriate for procedure C ($h_i + s_j^C - s_j^N$ is sufficiently positive) or very appropriate for procedure N ($h_i + s_j^C - s_j^N$ is sufficiently negative) then increasing physician information (precision $\bar{D}_j$) improves patient outcome.*

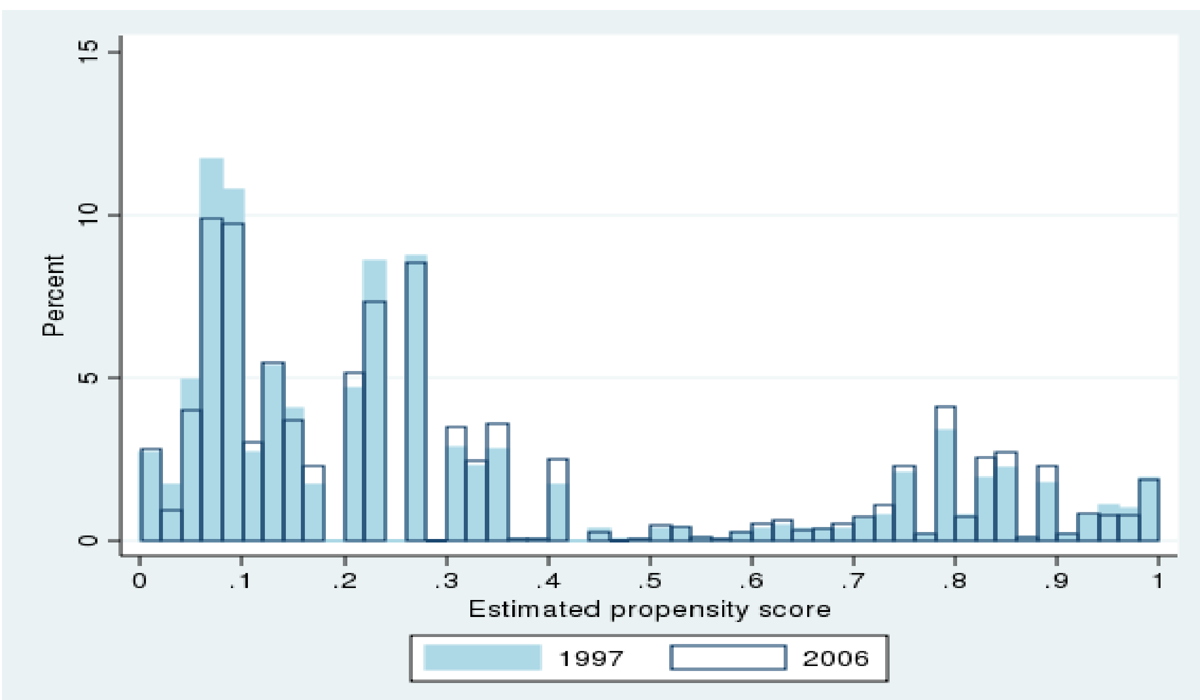
This proposition follows immediately from (12) and the fact that $F' > 0$. As in the case where the physician cannot perfectly observe h_i^r we have that in the extreme cases improved information makes matters better. Condition (12) gives the precise conditions under which improvements in information improve outcomes. Observe that if there are no pecuniary rewards to procedure choice ($m_j\left(\Delta P\right) = 0$) and h_j^0 then improvements in information always improve outcomes.

This analysis has a number of implications. First, it shows that changes in the quality of information that the physician has regarding a patient, but not her procedural skill, affect our measure of diagnostic skill. The difference now is with regards to the sign of the effect. In the case considered in this Appendix, our measure of diagnostic skill is *negatively* correlated with both physician information and health outcomes. This prediction is rejected by the data since we find that our measure of diagnostic skill is *positively* correlated with patient outcomes. Hence we can reject the hypothesis that variations in physician decision capabilities arise solely because of superior information relative to the information we have in h_i^r. If confirmed by future research this result has important practical implications. In particular it suggests that physician performance may be improved with the use of system wide data collected from a large sample of physicians.

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



Appendix Figure 2: Shift in Medical Risks over Time



Appendix Table 1: Models Estimated Using First Births Only

Medical Risk:	Very Low*	Low	Medium	High
C-Section				
Diagnostic Skill		-0.057	-0.016	0.145
		(0.017)	(0.014)	(0.013)
Procedural Skill Difference		0.116	0.071	0.130
		(0.034)	.036)	(0.044)
Price Differential (x 100)		0.513	0.674	0.084
		(0.103)	(0.104)	(0.088)
C-section Risk		0.518	0.950	0.692
		(0.071)	(0.016)	(0.063)
R-squared		0.023	0.057	0.025
Any Bad Outcome				
Diagnostic Skill		-0.020	-0.030	-0.032
		(0.009)	(0.008)	(0.017)
Procedural Skill Difference		-0.173	-0.223	-0.809
		(0.036)	(0.032)	(0.050)
Price Differential (x 100)		0.020	-0.060	-0.051
		(0.075)	(0.055)	(0.094)
C-section Risk		0.198	0.136	0.707
		(0.048)	(0.010)	(0.066)
R-squared		0.005	0.007	0.042
Neonatal Death				
Diagnostic Skill		-0.003	-0.005	-0.024
		(0.001)	(0.001)	(0.004)
Procedural Skill Difference		-0.006	-0.011	-0.065
		(0.003)	(0.003)	(0.016)
Price Differential (x 100)		0.009	0.027	-0.018
		(0.007)	(0.006)	(0.025)
C-section Risk		-0.024	0.022	-0.100
		(0.014)	(0.003)	(0.024)
R-squared		0.004	0.005	0.025
# Observations	1691	91802	258083	33491

Notes: Standard errors are clustered on the physician and shown in parentheses. Regressions also included all of the variables listed in Table 4.
* Since there were only 1691 first births in the lowest risk category we do not show estimates for this category.

Appendix Table 2: Effect of Alternative Diagnostic Skill Measure

Medical Risk:	Very Low	Low	Medium	High
C-Section				
Diagnostic Skill	-0.068	0.032	0.184	0.558
	(0.018)	(0.016)	(0.035)	(0.031)
Procedural Skill Difference	0.020	0.030	0.071	0.113
	(0.017)	(0.021)	(0.038)	(0.029)
Price Differential (x 100)	0.135	0.404	0.538	-0.008
	(0.051)	(0.063)	(0.094)	(0.054)
C-section Risk	0.182	1.074	0.869	0.778
	(0.090)	(0.037)	(0.010)	(0.020)
R-squared	0.015	0.041	0.219	0.067
Any Bad Outcome				
Diagnostic Skill	-0.036	-0.024	-0.064	-0.090
	(0.010)	(0.009)	(0.016)	(0.021)
Procedural Skill Difference	0.027	-0.147	-0.296	-0.779
	(0.025)	(0.021)	(0.030)	(0.032)
Price Differential (x 100)	0.090	0.044	0.003	0.028
	(0.053)	(0.041)	(0.052)	(0.054)
C-section Risk	0.068	0.295	-0.038	0.380
	(0.068)	(0.022)	(0.006)	(0.019)
R-squared	0.006	0.013	0.010	0.053
Neonatal Death				
Diagnostic Skill	-0.008	-0.002	-0.012	-0.035
	(0.003)	(0.001)	(0.003)	(0.005)
Procedural Skill Difference	0.005	-0.010	-0.021	-0.063
	(0.004)	(0.002)	(0.004)	(0.008)
Price Differential (x 100)	0.015	0.009	0.027	-0.014
	(0.007)	(0.004)	(0.007)	(0.012)
C-section Risk	-0.109	0.014	-0.013	0.048
	(0.026)	(0.006)	(0.002)	(0.007)
R-squared	0.004	0.004	0.008	0.016
# Observations	104902	364268	381745	117833

Notes: Standard errors are clustered on the physician and shown in parentheses. Regressions also included all of the variables listed in Table 4. The alternative diagnostic measure is the share of high risk patients who receive a C-section less the share low risk who receive a C-section. The mean is .825 and the difference between 75th and 25th p'tile is .086. The coefficient of -.068 in the first column implies that a change of this magnitude would reduce C-sections by -.068*.086=-.006 pp on a baseline of .06.

Appendix Table 3: Calculating Percent Changes

	C-Section				Bad Outcomes				Neonatal Death			
	V. Low	Low	Med	High	V. Low	Low	Med	High	V. Low	Low	Med	High
Base rate of outcome	0.06	0.115	0.439	0.891	0.038	0.048	0.085	0.086	0.00322	0.00238	0.00478	0.00785
Physician Level Estimates												
Coeff on Diagnostic Skill	-0.044	-0.057	0.077	0.194	-0.030	-0.022	-0.042	-0.033	-0.005	-0.003	-0.010	-0.017
Coeff on Procedural Skill	0.036	0.071	0.055	0.072	0.040	-0.137	-0.280	-0.772	0.007	-0.008	-0.016	-0.057
Coeff on Price (*100)	0.125	0.389	0.537	-0.018	0.082	0.040	-0.003	0.029	0.013	0.008	0.025	-0.014
%ch w 25-75 Diagnostic Skill	-0.158	-0.107	0.038	0.047	-0.170	-0.099	-0.106	-0.083	-0.334	-0.271	-0.450	-0.466
%ch w 25-75 Procedural Skill	0.037	0.038	0.008	0.005	0.065	-0.177	-0.204	-0.557	0.135	-0.208	-0.208	-0.450
%ch w $2600 Price Increase	0.054	0.088	0.032	-0.001	0.056	0.022	-0.001	0.009	0.105	0.087	0.136	-0.046
Market Level measures												
Coeff on Diagnostic Skill	-0.166	-0.157	0.062	0.542	-0.103	-0.116	-0.102	-0.122	-0.006	0.000	-0.007	-0.014
Coeff on Procedural Skill	0.111	0.054	0.207	0.248	0.08	-0.124	-0.181	-0.543	0.019	0.006	0.003	-0.047
Coeff on Price (*100)	0.116	0.311	0.300	-0.158	0.093	0.060	0.132	0.163	0.011	0.009	0.032	-0.006
%ch w 25-75 Diagnostic Skill	-0.191	-0.094	0.010	0.042	-0.187	-0.167	-0.083	-0.098	-0.128	0.000	-0.101	-0.123
%ch w 25-75 Procedural Skill	0.054	0.014	0.014	0.008	0.061	-0.075	-0.062	-0.183	0.171	0.073	0.018	-0.174
%ch w $2600 Price Increase	0.050	0.070	0.018	-0.005	0.064	0.033	0.040	0.049	0.089	0.098	0.174	-0.020

Notes: The difference between the 75th and 25th percentiles of diagnostic skill are .215 and .069 for the physician-level and market-level measures respectively. The corresponding differences for procedural skills are .062 and .029.
Coefficients for physician-level measures come from Tables 4 and 5. Coefficients for market-level measures come from Table 8.