Specialization and Sorting in the Obstetrics Market

Andrew Epstein
Yale University

Jonathan Ketcham
Arizona State University

Sean Nicholson
Cornell University

April 2006

Abstract:

The welfare implications of variations in how physicians treat patients depend on whether patients have different optimal treatments and are treated by physicians who are likely to provide their optimal treatment. We examine the extent to which expecting mothers direct themselves, or are directed to, particular physicians based on patients' preferences for physicians' treatment styles and patients' health conditions. We capitalize on the largely random assignment of weekend patients to physicians due to call schedules and the concentration of induced deliveries and scheduled c-sections, opportunities for a patient to choose her physician, to weekdays. Using Florida and New York discharge data from 1999 to 2004 linked to information on physician practices, we find that one-third of the variation in treatment styles across physicians is due to patient-physician matching on unobserved characteristics, which implies that a considerable part of the variation in medical treatment rates may enhance welfare. In one-third of the group practices, certain physicians specialize in treating weekday patients with relatively high observed risk and others with relatively low observed risk, and there is more than twice as much variation across physicians within a practice in patients' observed health than one would expect if weekday patients were randomly assigned.

I. Introduction

Extensive research has documented that there are substantial variations in the amount and type of medical care people receive across regions (Wennberg and Gittelsohn, 1973; Chassin et al., 1986; Wennberg, Freeman, and Culp, 1987; Phelps and Parente, 1990; Wennberg, Fisher, and Skinner, 2002) as well as across physicians within a region (Roos et al., 1986; Stano and Folland, 1988; Goyert et al., 1989; Welch, Miller, and Welch, 1994; Grytten and Sorensen, 2002; Epstein and Nicholson, 2005; Grant, 2004).[1] Most of these studies interpret variations across regions and physicians as reducing welfare.[2] Phelps and Parente (1990), for example, estimate there is a $33 billion annual welfare loss due to variations in hospital use rates under the assumption that the mean observed use rate is the optimal rate.

If consumers have homogeneous optimal treatments, then welfare will be reduced when physician treatment styles vary from the benchmark. In many circumstances, however, patients are likely to have different optimal treatments based on their preferences and health. If so, then variations in treatment styles, or product variety, will increase welfare if consumers are able to search successfully for physicians who are likely to provide their optimal treatment. The objective of this paper is to examine the extent to which expecting mothers direct themselves, or are directed to, particular physicians. We examine patient-physician matching based on patients' preferences and health conditions, and physicians' treatment styles. Treatment styles are defined as the probability a physician would perform a cesarean section versus a vaginal delivery, controlling for the patient's observed health and the status of the pregnancy.

---

[1] Treatment use rates could vary across regions due to differences in prices, income, patients' health status, patients' preferences, or physicians' ability or willingness to induce demand for their services. Phelps and Mooney (1993) conclude that these factors collectively explain very little of the differences in the amount of medical care received. Several other authors concur with this assessment, including Chasin (1986), Bikhchandani et al. (2001) and Newhouse (2002).

[2] An exception is a paper by Chandra and Staiger (2004) in which providers develop competencies in certain treatment styles. Patients who are good candidates for surgery will have better outcomes if they live in an area that specializes in surgery versus an area that specializes in medical treatment, for example, and likewise for patients who are good candidates for medical treatment.

Several features of the market for obstetrics facilitate empirical research on the welfare implications of product differentiation. First, most full-time obstetricians perform more than 100 deliveries per year, which allows us to measure their treatment style fairly precisely. Second, administrative hospital discharge data can be used to determine physician treatment styles because over 90 percent of deliveries occur at hospitals. Third, pregnant women have time to search for physicians who closely match their preferences (e.g., a physician with a low propensity to perform a cesarean section) or may specialize in treating pregnant women with their health condition (e.g., severe hypertension, diabetes, a fetus in the breech position).

Fourth, most induced deliveries and scheduled c-sections -- deliveries in which a patient can explicitly choose a particular physician within the obstetrical practice or the practice can assign a patient to the physician who specializes in her condition -- occur on weekdays when hospitals are fully staffed and when many physicians prefer working. By contrast, women who go into labor on a weekend usually match with the physician in the practice who is on call that day. Therefore, physician treatment styles and patient characteristics (e.g., illness severity of a physician's patients) based on weekday deliveries should include the effect of patient sorting, whereas measures based on weekend deliveries should reflect how each physician would treat a random sample of patients. If there is greater variation across physicians in their weekday versus their weekend treatment styles, this would provide evidence that patients are choosing or being directed to specialists within a group practice.

Public and private health insurers are increasingly collecting information on how physicians and hospitals treat patients, and the patients' health outcomes resulting from those treatments. The Centers for Medicare & Medicaid Services (CMS), for example, now posts on the internet how well each U.S. hospital treats heart attack patient (e.g., the percentage of patients who receive a beta blocker drug), congestive heart failure patients, and pneumonia patients. New York and Pennsylvania publish the risk-adjusted morality of each cardiac surgeon's patients. This information is disseminated to the public or is used to adjust physician and hospital

payments.[3] The information may help consumers find the physician or hospital that is likely to provide their optimal treatment. However, the information may also encourage outlier physicians and hospitals to alter their treatment styles to be closer to the market average, thereby reducing the amount of product differentiation. This paper addresses whether consumers are currently able to match with physicians based on their preferences and health characteristics, how matching occurs, and whether matching affects women's health outcomes.

Using Florida and New York discharge data from 1999 to 2004 linked to data on physician practice characteristics, we measure the magnitude of sorting on unobserved characteristics -- preferences and unmeasured health -- by comparing the amount of variation across physicians in their treatment styles and patient characteristics on the weekend to the amount on weekdays. We find that the variance in physicians' weekend risk-adjusted c-section rates is 66 percent as large as the weekday variance, which implies that up to one-third of the weekday variation may be due to patient-physician matching based on preferences for a physician's treatment style. Therefore, a considerable component of the variation could be welfare enhancing.

We test for different sorting mechanisms by examining correlations between a physician's weekend treatment style relative to the colleagues in his group practice, and the number of deliveries, the type of deliveries, and the types of patients he treats during the week relative to his practice's average. Physicians with high weekend c-section rates perform a relatively large number of scheduled weekday patient-choice (non high-risk) c-sections, which is consistent with women seeking out physicians who share their views regarding the net benefits of a c-section.

To examine sorting based on patients' observed clinical risk factors, we compare the actual distribution of weekday deliveries of physicians within a practice to the distribution one

[3] Mark McClellan, the administrator of CMS, forecast in 2004 that 20 percent of Medicare's payments would be based on physicians' and hospitals' performance or outcomes within five years.

would expect if patients were randomly assigned to physicians in that practice. In one-third of the group practices, certain physicians specialize in treating high-risk patients and others in low-risk patients, as measured by the patients' observed health. There is more than twice as much variation across physicians within a practice than one would expect if weekday patients were randomly assigned. Although we find no evidence that high-risk patients are being matched with physicians who have a high propensity to perform c-sections, other methods of matching could improve health outcomes.

Finally, we exploit the largely random assignment of women who go into labor on the weekend to a physician in the group practice. We instrument for whether a patient receives a c-section with her physician's propensity to perform c-sections on weekend patients in order to estimates the causal effect of a c-section on a woman's health. We cannot reject the hypothesis that performing more c-sections on "marginal" patients -- women who would receive c-sections if treated by physicians in their practices with favorable views of c-sections, but would not receive c-sections if physicians with less favorable views were on-call instead -- would have no effect on the mother's health outcome, which is consistent with physicians using c-sections appropriately.

II. The Market for Obstetrical Services

Our measure of an obstetrician's practice style is his risk-adjusted c-section rate -- the probability a patient would receive a c-section if treated by the physician, controlling for observed measures of the patient's health and the status of the pregnancy. A c-section is a surgical procedure and, as such, is more expensive than a vaginal delivery, requires a longer recovery period, and presents some health risks for the mother. In certain clinical situations, however, a c-section clearly improves the expected health of the infant and/or the mother.

The c-section rate in Florida has increased sharply since 1996 and now stands at 33.1 percent of all deliveries, as displayed in Figure 1. C-sections can either be scheduled in advance of labor or can occur after a woman goes into labor. The recent increase in the Florida c-section

rate has been driven primarily by a doubling of the scheduled c-section rate, from 11.0 percent in 1996 to 23.4 percent of all deliveries in 2003. Scheduled c-sections now account for about two-thirds of all c-sections in Florida. Figure 1 also shows the percentage of total deliveries that are scheduled c-sections separately for women with and without a high-risk condition.[4] The latter, sometimes referred to as "patient-choice" c-sections, now constitute over one-half of all scheduled c-sections in Florida.

Epstein and Nicholson (2005) show that two women with the same characteristics can have a substantially different probability of receiving a c-section based on the obstetricians they choose. Specifically, the standard deviation of the within-market variation in physicians' risk-adjusted c-section rates is 6.0 percentage points. If patients have homogeneous preferences for a physician's treatment style and there is a single style that produces the best expected outcome for all types of patients, then this variation reduces welfare. However, it seems plausible that some women may prefer an obstetrician with a high propensity to perform a c-section while others may prefer a more conservative obstetrician, depending on the mother's health, the fetus' health, the status of the pregnancy, and how the mother interprets the risks associated with a vaginal delivery versus a c-section. Therefore, some of the observed variation in physicians' treatment styles may be caused by women with different preferences matching with physicians with particular styles.

There are other types of patient-physician matching that could improve welfare relative to randomly paired patients and physicians. Physicians may develop expertise in treating particular types of patients, perhaps based on where they trained as a resident. Consider a situation where a woman's obstetrician discovers late in the pregnancy that the fetus is not in the vertex position (i.e., head first), a condition that occurs in about seven percent of pregnancies. After discussing with the patient whether she wants to attempt a vaginal delivery, they decide to schedule a c-section with another member of the group practice who is believed to be highly skilled in these

[4] We consider a woman to be "high-risk" if she has one or more of the first 12 health conditions listed in Table 1.

types of deliveries. Below we develop specific hypotheses to test whether patients and physicians are matching based on patient preferences for physicians' treatment styles and the mother's observed clinical risk factors.

III. Data

We construct our sample from the Florida and New York hospital discharge data sets for the 1999 to 2004 time period. The data sets contain information on 757,000 deliveries that occurred at every non-federal, short-term acute care hospital in those states. We observe the mother's demographic information (age, race), her insurance coverage (e.g., HMO, Medicaid), codes for the primary diagnosis and secondary diagnoses, procedure codes that allow us to determine whether the baby was delivered vaginally or via c-section, a unique and consistent (across hospitals and years) physician identifier, a unique and consistent hospital identifier, and the quarter and year the patient was admitted. The diagnoses codes allow us to control for objective health conditions that affect the probability a physician will perform a c-section (e.g., whether a woman had a c-section prior to this delivery, whether the fetus was malpositioned during the delivery such as in the breech position, or whether the labor occurred before the fetus was full-term). Women are much more likely to receive a c-section in these situations. To minimize concerns about measuring physician treatment styles and birth outcomes with error, we restrict the analysis to physicians who performed at least 100 weekend deliveries between 1999 and 2004 on women who went into labor.

Henry *et al.* (1995) and Gregory *et al.* (2002) developed a method to determine whether or not a woman labored. Women who delivered vaginally or had codes indicating fetal distress, labor abnormalities, cord prolapse, or a breech converted to vertex presentation were interpreted has having labored. We use a coding system established by Health Grades (HealthGrades, 2004) to determine whether or not a mother experienced an adverse outcome during the delivery. Examples include injury to pelvic organs, rupture of the uterus, 4^{th} degree perineal laceration,

anemia due to acute loss of blood, or major puerperal infection. In 1999, 10.9 percent of vaginal deliveries and 8.6 percent of c-sections resulted in an adverse outcome for the mother using the Health Grades definition.

This definition, however, does not allow for the possibility that the utility loss may vary across the 66 different discrete adverse outcomes for a c-section, and the 64 discrete adverse outcomes for a vaginal delivery. Therefore, we create an alternative outcome variable by regressing the number of days that a woman spends in the hospital, separately for the women who delivered vaginally or by c-section, on indicator variables for whether the woman experienced each of the adverse events specified by Health Grades and hospital fixed effects. Our key assumption is that the welfare loss from an adverse event is proportional to the marginal effect of that event on a mother's hospital stay, controlling for the hospital of admission.

We estimate the regressions separately for all vaginal and c-section deliveries in 1998, the year prior to the beginning of our sample. Regression results for Florida are reported in the appendix. In the c-section regression 20 of the events have a positive coefficient that is greater than 0.50 and significantly different from zero, whereas 11 events in the vaginal delivery regression have a positive coefficient of that magnitude. We multiply the coefficient estimates on each outcome variable by a mother's actual birth experience to generate a continuous measure of her health outcome, with larger values indicating a worse outcome. We also create a dichotomous outcome variable that takes on the value of one if a woman experienced any of the "major" adverse outcomes, where "major" is defined as an outcome with an incremental length of stay of 0.50 or greater.

Because the physician identifiers are consistent and the data include all hospital discharges, we are able to examine a physician's entire inpatient practice. Using the physician license numbers as a link, we obtained each physician's office telephone number in April 2005 from the Florida State Licensure Database and American Medical Information, a private company that tracks physicians for general marketing purposes. We used the telephone number to identify

physicians who practiced by themselves, and to identify the members of each group practice. Our analysis is restricted, for reasons we elaborate on below, to physicians who practice in a group. Of the 733 physicians who performed 100 or more weekend laboring deliveries in Florida between 1999 and 2004, 342 are identified as members of a group practice.[5] There are 431 physicians in New York practicing in a group who exceed our delivery thresholds, for a total of 773 physicians included across both states.

Sample means and standard deviations for the patient-level data set for are reported in Table 1, separately for weekend deliveries where the mother went into labor and all weekday deliveries.

IV. Methods

The objective of this paper is to identify whether patients match with physicians, how they match, and the implications of matching on health outcomes. We test matching based on both patient characteristics that we cannot observe, including patient preferences, and observable clinical factors. We employ different strategies to test matching based on observables versus unobservables, although both approaches rely on the random assignment of weekend patients to physicians within the group do define each physician's treatment style apart from patient selection. Because we rely on random assignment due to practice's call schedules, we focus on patients who have already selected a group practice. As a result, our analysis only addresses sorting of patients *within* a practice but does not consider sorting *between* practices. To test matching based on unobservable patient characteristics, we compare weekend deliveries where the mother went into labor, where we expect very little matching, to all weekday deliveries, where we expect the greatest amount of matching. To test matching based on observable patient characteristics, we compare the actual distribution of weekday deliveries to physicians within a

[5] The practice affiliation of seven physicians could not be identified because they were no longer practicing in Florida.

practice to the distribution one would expect if patients were randomly assigned to physicians in that practice. We then examine whether observed matching appears to be explained by the physicians' treatment styles.

In most obstetrical group practices, the on-call physician(s) performs the deliveries of all the patients from the practice that go into labor on the weekend. Therefore, most weekend patients are randomly assigned to a particular member of the practice. There are likely to be some exceptions to this general rule. A small number of scheduled c-sections and induced deliveries occur on the weekend -- situations where the patient and the physician are matched. We restrict the weekend sample to deliveries where a woman goes into labor, thereby excluding scheduled weekend c-sections and reducing the amount of non-random matching in this sample. Unfortunately, we cannot observe from the discharge data whether or not a delivery was induced. Twenty percent of deliveries in the United States in 2000 were induced.[6] In 1997, approximately 15 percent of the induced deliveries occurred on the weekend, so a small part of the weekend laboring sample will consist of patients and physicians who consciously matched.[7] As discussed below, we empirically test the assumption that weekend patients are randomly assigned within the practice, and we repeat the analyses excluding the groups that violate this assumption.

Women who deliver on a weekday are much more likely to choose their physician, or to be matched by the practice to a particular physician, than women who go into labor on the weekend. As on the weekend, most obstetrical group practices designate a physician to perform deliveries on all patients from the practice that go into labor during a particular 24 hour period.[8] If a patient and/or the physicians in the practice want to guarantee that a particular physician will deliver a woman's baby, they will either schedule labor to be induced or schedule a c-section

---

[6] National Vital Statistics Report, 2003: Births: Final Data for 2002.

[7] National Vital Statistics Report, 1999, Trends in the Attendant, Place, and Timing of Births, and in the Use of Obstetric Interventions: United States, 1989-97.

[8] In some practices each physician will deliver their own patients who go into labor between 7am and 5pm, and then one member of the group will be on-call from 5pm to 7am to perform deliveries on patients who go into labor at night. This structure will increase the amount of patient-physician matching that occurs for weekday deliveries.

when the preferred physician is on-call at the hospital. Because hospitals are staffed more heavily during the week than the weekend and many physicians value leisure time on the weekend more highly than leisure time during the week, induced deliveries and scheduled c-sections are much more likely to occur on a weekday than a weekend. Twenty-one percent of weekday deliveries in our data set were scheduled c-sections (Table 1), and these are almost equally divided between patient-choice and high-risk c-sections. As mentioned earlier, 20 percent of all deliveries in the U.S. are induced, and these are 2.3 times more likely to occur on a weekday than a weekend. Therefore, physicians and patients consciously match for a substantial percentage of weekday deliveries.

Consider the following ordinary least squares regressions run on the sample of patients who went into labor and delivered on the weekend to a physician who is a member of a group practice:

$$(1)\quad \text{C-section}_i = \beta_0 + \boldsymbol{\beta_1}\mathbf{MD} + \boldsymbol{\beta_2}\mathbf{X}_i + \boldsymbol{\beta_3}\mathbf{Year} + \varepsilon_1$$

C-section is an indicator variable for whether a woman received a c-section, **MD** is a vector of physician fixed effects, **X** is a vector of patient characteristics, and **Year** is a vector of fixed effects for year of delivery. We include patient characteristics such as the patient's age, type of health insurance, and health conditions (e.g., whether the woman had a c-section in a prior delivery, multiple gestation, premature labor, antepartum bleeding) to control for medical factors that affect the likelihood of receiving a c-section and the outcome of the delivery. This allows us to measure physicians' treatment styles and outcomes if they were to treat an identical (based on observed characteristics) patient population. Specifically, the coefficients on the physician indicators ($\boldsymbol{\beta_1}$) measure the probability that a particular physician will perform a c-section, controlling for observed patient characteristics.

If patients who deliver on a weekend are essentially randomly assigned to the physician within the practice who happens to be working at the hospital that day, then a physician's

weekend c-section rate will be a clean measure of his treatment style.[9] The variation between physicians in risk-adjusted weekend c-section rates should be due to differences in their perceptions about how to treat a common set of patients, not due to physician-patient matching within a practice based on the patient's preferred treatment or the patient's unmeasured health.

We will estimate an identical regression for all weekday deliveries (laboring and non-laboring deliveries) performed by physicians in group practices:

(2) $\text{C-section}_i = \beta^d_0 + \mathbf{\beta^d_1 MD + \beta^d_2 X_i + \beta^d_3 Year} + \varepsilon_2$

The superscript "d" refers to a weekday delivery. We separately test for patient sorting based on patients' observed clinical characteristics and unobserved characteristics -- preferences for physicians' treatment styles and unobserved health factors.

One possibility is that some patients observe physicians' treatment styles and choose particular physicians whose styles are consistent with their preferences, or patients are directed to such physicians by other members of the practice or primary care physicians. For example, a woman who places a high cost on the adverse outcomes associated with a c-section could reduce the probability she will have a c-section by scheduling her labor to be induced on a day when a "conservative" obstetrician is on call, or she may request that a particular physician perform the delivery if the patient goes into labor during the week. Conversely, consider a woman without a high risk health condition who wants to avoid laboring altogether and asks her physician to perform a scheduled, patient-choice c-section. If the physician refuses her request, she may be referred to another member of the practice who is willing to perform such procedures. An obstetrician quoted in a 2003 *New York Times* article confirms that these types of decisions occur: "Obstetrics has become very consumer-driven. When a woman can't get what she wants from one doctor, she'll go to another."

[9] For larger groups, multiple physicians might be on call, but most patients are likely assigned based on the timing of their onset of labor rather than consciously sorted between on-call physicians. Another possibility that we test is whether a group treats patients in multiple hospitals and there is non-random assignment of the group's physicians and patients across hospitals.

If patients are sorting on physician styles as described above, there should be more variation across physicians in their weekday c-section rates than in their weekend rates: variance(β^d_1) > variance(β_1). This should be true even when c-section rates are adjusted for observed patient characteristics, because patients with unobserved (to us) preferences will be matching with specific physicians -- patients who want intensive treatment match with physicians with a large β_1 value, and patients who want less intensive treatment match with physicians with a small β_1 value. A physician who performs many patient-choice scheduled c-sections, for example, will have a large positive β^d_1 because his actual proportion of c-sections will exceed the predicted proportion based on his patients' observed characteristics, and vice versa for physicians who are attracting patients who want to be treated conservatively. We would also expect variance(β^d_1) to exceed variance(β_1) if patients who have poor unobserved (to us) health match with physicians who have a high propensity to perform c-sections, presumably in order to improve their expected health outcome.

The test described above is a joint test that patients with particular preferences and/or unobserved health choose their physicians for weekday deliveries *and* they choose their physicians based on the physicians' c-section treatment styles. If instead patients choose physicians based on something that is not strongly correlated with the c-section rate, such as birth outcomes, then we may not find greater variation across physicians during the week than on the weekend even though sorting is occurring.

One shortcoming of this test is that it does not distinguish between physician-driven and patient-driven matching. Physicians with a favorable view of c-sections, as defined by their high weekend c-section rate, may convince their "borderline" patients to schedule a c-section with them during the week. If this behavior is the prime driver of the greater variation in weekday than weekend treatment styles, it has different welfare consequences than if patients are initiating or participating fully in the treatment decisions. We therefore also test whether correlation($\beta_{j1} - \beta_p$, $Q^d_j - Q^d_p$) > 0, where ($\beta_{j1} - \beta_p$) is the difference between physician j's and his practice's

average weekend risk-adjusted c-section rate, and ($Q^d_j - Q^d_p$) is the difference between the number of weekday scheduled patient-choice c-sections physician j performs relative to the practice average. That is, do physicians who perform a large number of c-sections relative to their colleagues on randomly assigned patients also perform a large number of scheduled weekday patient-choice c-sections, or a large number of scheduled weekday high-risk c-sections? The results for patient-choice deliveries addresses the extent to which patient preferences are driving the sorting, because our assumption is that patient-choice c-sections are initiated by the patient whereas a scheduled high-risk c-section could be initiated by either party.

A second form of sorting could be based on the patient's observed clinical condition and the comparative advantage of physicians within the practice. If a group practice has a physician who specializes in certain high-risk deliveries, the practice might schedule many of these as c-sections on a weekday when that specialist is on-call at the hospital. We explore whether this type of specialization and sorting occurs by comparing the actual characteristics of weekday deliveries among physicians within a practice to the pattern one would expect to observe if weekday patients were randomly assigned to a physician within that practice. Specifically, we perform 1,000 iterations of a simulation. In each iteration a weekday patient is randomly assigned to a physician in the patient's practice such that each physician receives his actual number of weekday patients. After each iteration, we record the mean patient characteristics for each physician (e.g., the proportion of his patients who have a high-risk condition or the proportion who are over 35 years of age), as well as the variance of these characteristics across physicians within a practice. If the actual variance across physicians within a practice exceeds the 95th percentile of the simulated variances for that practice, we conclude that at least one of the physicians in this practice is specializing in that patient characteristic.[10]

[10] We could also test for specialization in patients' observed clinical risk factors by comparing the variation across physicians in the average characteristics of their weekday patients relative to their weekend patients, much as we do to test sorting on unobservables. Although there is substantially more variation in patient characteristics across physicians during the week than on the weekend, this could be due in large part to the

One shortcoming of the specialization test outlined above is that it does not distinguish between a situation where physicians in a practice attract different types of patients and perform all the deliveries on their patients who deliver during the week, and a situation where patients are reassigned to physicians within the practice based on the physicians' comparative advantages and information the practice learns about the woman's health during the pregnancy. We try to distinguish between these two explanations by examining whether practices specialize in patient characteristics that cannot be forecast when a woman initially selects "her" physician in the practice, such as whether her fetus will be in the vertex or breech position in the final trimester of the pregnancy.

These general tests do not prove that clinical sorting improves health outcomes, only that physicians and/or patients presumably believe that it does. We also test for a more specific type of clinical sorting: are patients who are at high risk of needing a c-section directed to physicians with favorable views about the benefits of c-sections? Specifically, we test whether correlation$(\beta_{j1} - \beta_p, \hat{C}^d_j - \hat{C}^d_p) > 0$. Are patients who are observed to be at high risk of needing a c-section directed to physicians in the practice with a favorable view of c-sections? $\hat{C}^d_j$ and $\hat{C}^d_p$ are the mean predicted c-section probability of physician j's weekday patients and the weekday patients of physician j's practice, respectively, where the predicted rate is derived using all coefficients from (2) except β^d_1.

In addition to documenting whether and how patients sort across physicians within a group, we seek to quantify the impact of sorting on outcomes. The largely random assignment of weekend patients to physicians in group practices offers a rare opportunity to examine the health production function. Are a woman's outcomes better or worse if she goes into labor on a weekend when a physician with a high risk-adjusted c-section rate, relative to the average c-section rate in the practice, is on call? If so, then if women who are appropriate for a c-section

---

fact that weekend patients are much more homogeneous in terms of clinical risk factors than weekday patients.

are matching with physicians with a propensity to perform c-sections on weekdays, outcomes should improve as a result.

This section adapts a framework developed by Chandra and Staiger (2004), which examines whether heart attack patients who are most appropriate for intensive treatment (e.g., receiving a catheterization treatment) have better outcomes in regions that treat heart attack patients intensively, and whether heart attack patients who are most appropriate for medical treatment have worse outcomes in such regions. Among women who go into labor and deliver on the weekend to a group physician, our equation of interest is:

$$(3)\quad \text{Outcome}_i = \sigma_0 + \sigma_1 \text{C-section}_i + \boldsymbol{\sigma_2 X_i}, + \boldsymbol{\sigma_3 Year} + \boldsymbol{\sigma_4 Practice} + \varepsilon_3,$$

where "Practice" is an indicator variable for each of the 189 group practices. The main problem with interpreting σ_1 as a causal effect of c-section is that once a woman is in labor, the physician decides whether or not to perform a c-section based on variables we observe plus factors we do not observe. If women who receive c-sections are in worse health than they appear based on **X**, which is likely, σ_1 will be biased in the direction of finding lower returns to a c-section.

We address this concern by instrumenting for C-section_i with a slightly modified version of physician j's risk-adjusted weekend c-section rate among patients who go into labor:

$$(4)\quad \text{C-section}_i = \varphi_0 + \boldsymbol{\varphi_1 X_i} + \varphi_2 \text{OBdiff}_j + \boldsymbol{\varphi_3 Year} + \boldsymbol{\varphi_4 Practice} + \varepsilon_4$$

Following Duggan (2005), OBdiff is the difference between the proportion of physician j's weekend patients who actually received a c-section and the proportion predicted to receive a c-section based on β_0, β_2, and β_3 from equation (1), with patient i removed from the calculation. It is essentially the physician's risk-adjusted treatment style (β_1), on all patients other than patient i. The coefficient φ_2 measures the change in the probability that a woman would receive a c-section if she went into labor on the weekend when a physician in her practice with a favorable view of the benefits of a c-section was on-call, versus a physician with a less favorable view. Therefore,

the coefficient σ_1 measures the health consequences of a c-section for "marginal" patients only – women who would receive a c-section if the former physician was on-call but not if the latter physician was on-call.

There are two important assumptions underlying the interpretation of σ_1 as the causal effect of c-section. First, variation in OBdiff between physicians within a practice is driven by differences in their perceptions regarding how an identical set of patients should be treated, rather than differences in unmeasured patient preferences or health or physician skill. If physicians that perform relatively more c-sections are also those that are better at c-sections, then the results would overstate the benefits of c-section. Second, we assume that once we control for patients' observed health, obstetricians only affect women's outcomes through their decision of whether or not to perform a c-section. If physicians who perform a lot of c-sections are also better/worse along a number of other dimensions, such as providing effective prenatal care or benefiting from learning-by-doing, then σ_1 could be biased. We estimate equations (3) and (4) using two-stage least squares for all weekend patients.

V. Results

In Table 2 we report coefficient estimates on the patient characteristics from equations (1) and (2). The sample in the first specification consists of all weekend laboring deliveries performed by a member of a group practice, and in the second specification consists of all weekday deliveries performed by the same physicians. A total of 773 physicians performed 100 or more weekend deliveries between 1999 and 2004 and were thus included in the sample. Focusing on physicians with relatively high patient volumes should reduce measurement error in the practice style coefficients. On average, each physician performed 791 weekday and 188 weekend laboring deliveries over the six-year period. After performing these two regressions, we recover the physician-level weekend and weekday risk-adjusted c-section rates (the $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta^d_1}$ coefficients on the physician indicator variables) for each physician, and estimate the predicted

probability that each patient would receive a c-section based on her observed characteristics (but not $\boldsymbol{\beta_1}$ or $\boldsymbol{\beta^d_1}$).

Results of the simulation are reported in Table 3. $\hat{C}_j$, the mean predicted c-section probability of physician j's weekend patients based on their observed characteristics, is a summary measure of the clinical risk of physician j's patients. Most of the 189 group practices in Florida and New York appear to randomly assign their weekend patients to physicians in the practice according to the patients' clinical risk. The actual within-practice variance of $\hat{C}_j$ is significantly larger than the median variance in the 1,000 iterations for 23 percent of the practices.[11] The results are similar when examining the proportion of weekend patients who have a high-risk condition (one or more of the 12 conditions listed in Table 2).

Practices that do not randomly assign their weekend patients according to the patients' health are relatively large and have a relatively high proportion of poor-paying patients (Medicaid, uninsured, and patients covered by other public programs). As reported in the top panel of Table 4, practices where the physicians specialize clinically on the weekend have eight more physicians and a 15 percentage point higher mix of poor-paying patients, on average, than practices that do not specialize on the weekend. The most likely explanation is that many large practices provide on-call coverage at two or more hospitals, the patient populations differ between the hospitals, and certain physicians within the practice are scheduled exclusively or predominantly at one hospital. In eighty-three percent of practices with two to four physicians, every physician performs the predominance of their weekend deliveries at the same hospital, versus 27 percent of practices with five or more physicians. Some large practices may effectively function, therefore, as two (or more) sub-practices, where patients in one sub-practice are unlikely to be directed to physicians in another. For most of the remaining analyses we present results separately for all practices and only those practices that randomly assign their weekend patients.

[11] We test for significance at the 5-percent level using a one-tailed test. Specifically, if the actual within-practice variance exceeds the 95th percentile of the simulated variances for a particular practice, we conclude that the actual variance is significantly larger than the simulated variance.

In the second panel of Table 3 we report the percentage of practices that specialize during the week in a particular patient characteristic, among the 147 practices that randomly assign their weekend patients. If all practices actually randomly assign their weekday patients, we would expect the actual within-practice variance to exceed the median of the simulated variances for five percent of the practices. In fact, about one-third of the practices specialize according to a patients' observed clinical risk, as measured by the mean propensity a physician's patients will receive a c-section ($\hat{C}^d_j$), the proportion of a physician's patients who have at least one high-risk health condition, or the proportion of a physician's patients who are over 35 years of age. We use the ratio of the actual within-practice (across physicians) variance to the median of the simulated variance to measure the degree of a practice's specialization. The median of this ratio for $\hat{C}^d_j$ across the 147 practices is 2.24, which indicates there is more than twice as much variation across physicians within a practice than one would expect if weekday patients were randomly assigned.

As mentioned earlier, if physicians attract different types of patients to a practice and perform most of the weekday deliveries for those patients, it would appear as if the practice specializes in observed patient characteristics. Although this could still be considered specialization, it is not based on any information the practice acquires about the woman's health condition during her pregnancy, or any information the woman acquires during her pregnancy regarding the comparative advantages of the physicians in the practice. If no patients are reassigned to physicians in the last trimester of a pregnancy, then we would not expect to see any practices specializing in characteristics that can not be forecast in the first trimester.

To the best of our knowledge, all women have an equal probability of having a malpositioned fetus.[12] From Table 2, a woman with a malpositioned fetus has a 0.49 percentage point higher probability of receiving a c-section relative to a woman with a fetus in the vertex position. Twenty percent of the practices appear to direct such deliveries disproportionately to

[12] Physicians may differ in their willingness to perform version procedures, where they attempt to convert a malpositioned fetus to the vertex position prior to a mother's due date.

specific physicians in the practice (Table 3), presumably to physicians who are skilled at performing vaginal births in such circumstances and/or physicians skilled at scheduled high-risk c-sections. The actual within-practice variance for malpositioned fetuses is 33 percent larger than one would expect from random assignment. Because this is one of many different types of matching that could occur, we conclude that 33 percent represents a lower-bound for the amount of within-practice variation that is due to physician specialization and patient-physician matching on observed patient characteristics.

Fifty-five percent of practices specialize according to a patient's race (Table 3), which is likely to limit the amount of clinical specialization that could otherwise occur. At the bottom of Table 4 we report correlations between the intensity of weekday specialization in various patient characteristics and delivery methods, as measured by the ratio of a practice's actual variation to the median of its simulated variance. Practices that specialize on a patient's propensity to receive a c-section tend to also specialize in malpositioned fetuses, and scheduled c-sections. Practices that specialize according to a woman's age also tend to specialize according to race, but neither of these types of specialization is highly correlated with specialization according to clinical risk.

At the bottom of Table 3 we present information on specialization by type of delivery. A patient's delivery method is affected by her characteristics, preferences, and a physician's recommendation/decision. Therefore, these figures do not convey specialization based on observed patient characteristics alone. Forty-two percent of practices specialize in patient-choice scheduled c-sections. Below we present evidence that this is due, at least in part, to patients matching with physicians in the practice willing to perform c-sections on apparently low-risk women.

The characteristics of practices that randomize weekend deliveries but specialize during the week are presented in the middle of Table 4. Practices that specialize according to a patient's propensity to receive a c-section, a woman's age, and for patient-choice scheduled c-sections have more physicians, on average, than practices that do not specialize on these dimensions. This

is similar to a finding in Garicano and Hubbard (2003) that lawyers in large markets are more specialized than lawyers in small markets. There is no difference in payer mix between the practices that do and do not specialize.

One possible method of clinical specialization would be to assign high-risk patients to physicians with a high propensity to perform c-sections, because these patients presumably benefit from intensive treatment. In Panel A of Table 6, however, we find no evidence that patients are sorting in this manner, where we define a high-risk patient as one with a high predicted probability of receiving a c-section. There is no correlation between an individual physician's weekend c-section rate relative to his practice's average and the observed risk of his weekday patients relative to his practice's average, which indicates that matching is based on an attribute other than a physician's risk-adjusted c-section rate.

We now examine whether patients match with physicians based on unobservables: patient preferences and/or unmeasured health. If patients match with physicians based on preferences for treatment styles, there should be more variation across physicians in their weekday risk-adjusted c-section rates than in their weekend risk-adjusted rates: variance(β^d_1) > variance(β_1). In order to focus on variation across physicians within each health market, we subtract the relevant regional mean risk-adjusted c-section rate from each physician's rate, which expresses each treatment style as a difference from the market average.[13]

In Figure 2 we plot the deviations of β^d_1 (y-axis) and β_1 (x-axis) from the market averages. The correlation between the two values is 0.54; physicians who perform more c-sections than their peers in the market on the weekend among laboring deliveries, controlling for patient characteristics, also tend to perform more c-sections than their market peers during the week on all deliveries. In Table 5 we confirm that there is more variation in β^d_1 than β_1. The weekend standard deviation is significantly smaller than the weekday standard deviation, which is

[13] In order to estimate the risk-adjusted c-section rate for each market, we include a full set of market indicators in equation (1) and equation (2) instead of a full set of physician indicator variables. There are eleven health care markets in Florida and eight in New York.

consistent with sorting based on preferences for this particular treatment style and/or sorting on unmeasured health. Specifically, the weekend variance is 63 percent as large as the weekday variance. One-third of the weekday variation in treatment style is due to patient-physician matching based on unobserved patient characteristics, which is similar to the magnitude of selection on patients' observed health.[14] Therefore, a considerable component of the variation in physician practice styles could be welfare enhancing.

The weekend treatment styles (β_1) of physicians in practices that do not randomly assign weekend patients may already incorporate unmeasured patient characteristics. In fact, the ratio of the weekend variance to the weekday variance in treatment styles is similar (0.66) among the subset of physicians in practices that randomize weekend deliveries (column 2) of Table 5.

This comparison of variances does not distinguish between physician-driven and patient-driven matching. We explore these two possibilities in Panel B of Table 6, where we report the correlation of the difference between a physician's and his group's average weekend risk-adjusted c-section rate and the difference between a physician's and his group's average annual quantity of weekday scheduled c-sections. The correlation is not significantly different from zero for scheduled high-risk c-sections, which could be initiated by a physician or a patient, but is significant and positive (0.16) for patient-choice c-sections. The latter are presumably initiated by the patient. Low-risk patients who want a c-section appear to match with a physician in the practice who has a relatively favorable view regarding the benefits of a c-section. This supports the interpretation that product variety can improve patients' welfare, and patient preferences create a substantial amount of the observed variation in obstetricians' treatment styles. The correlation is smaller (0.09) but still significant when we restrict the sample to physicians in practices that randomly assign weekend deliveries.

Physicians in the sample performed an average of 14.2 patient-choice c-sections per year. A physician who is at the 90th percentile of the difference between his own weekend c-section

[14] $0.63 = (0.0337)^2/(0.0425)^2$

rate and that of his practice is predicted to perform 1.1 more patient choice c-sections per year than his colleague who is at the 10th percentile of the distribution.[15] This represents a fairly small (7.7 percent) increase in the physician's annual quantity of patient-choice c-sections. However, patient-choice c-sections are just one example of how patients with particular preferences can match to physicians with particular treatment styles.

The largely random assignment of weekend patients to a physician allows us to examine the effect of a c-section on a woman's health outcome. Our method addresses whether women's birth outcomes could be improved if c-sections were performed on "marginal" weekend patients – a woman who would receive a c-section if treated by a physician in her practice with a favorable view of c-sections, but would not receive a c-section if a physician with a less favorable view was on-call instead.

In the first row of Table 7 we report the coefficient estimate on an indicator variable for whether a woman actually received a c-section when equation (3) is estimated by ordinary least squares. The dependent variable is the mother's health outcome and the sample in the first column consists of all patients who went into labor and delivered on the weekend to a physician in a group practice. The ordinary least squares coefficient (0.084) is positive and significant: c-sections are associated with worse health outcomes for the mother. This is not surprising because a physician chooses the treatment decision based on more information than we observe, and women are likely to receive a c-section when the delivery is not going well ($\varepsilon_3 > 0$).

The remainder of Table 7 contains results from the IV estimation of equation (3) and equation (4). The c-section indicator variable is instrumented with OBdiff, the difference between the proportion of physician j's weekend patients who actually received a c-section and the proportion predicted to receive a c-section based on β_0, β_2, and β_3 from equation (1), with patient i removed from the calculation to avoid concerns about reverse causality. One key

[15] This analysis is based on the sample of physicians in practices that randomly assign weekend deliveries to members of the practice.

assumption of the IV model is that variation in OBdiff between physicians within a market is driven by differences in their perceptions regarding how an identical set of patients should be treated, rather than differences in unmeasured health or patient preferences. Note from Table 3 that for most practices there is very little variation between physicians in the observed health of their weekend patients, as expressed by the predicted probability of a c-section. This does not prove that there is no variation in unmeasured health, but makes it less likely. We also perform the IV estimation on the sample of physicians from practices that appear to randomly assign weekend patients.

The first-stage coefficient on OBdiff in the first column (0.27) is positive and highly significant. A patient who is seen on the weekend by a physician who performs more c-sections than expected based on the observed characteristics of his other weekend patients, is substantially more likely to receive a c-section herself. The coefficient on the c-section indicator in the second stage (0.11) is larger than the OLS insignificant. We would expect the coefficient on the c-section indicator in the IV estimation to be smaller (less positive) than in the OLS regression if the instrument is uncorrelated with the patient's unobserved health. This could occur if some of the practices match weekend patients to physicians based on a woman's unobserved health, which would create a correlation between a physician's risk-adjusted weekend c-section rate and factors that affect his patients' delivery outcomes.

In the second column we repeat the IV estimation on the sample of patients in practices that appear to randomly assign weekend deliveries to physicians. The first stage coefficient is positive and significant, but only marginally so; the instrument is fairly weak. As expected, the coefficient on the c-section indicator in the second stage is less positive in the IV model than the OLS model, although it is still insignificant. Therefore we cannot reject the hypothesis that c-

sections have little effect on the mother's health outcome at the margin, which is consistent with physicians using c-sections appropriately.[16]

VII. Conclusions

In this paper we explored the implication of product differentiation and patients' preferences in the market for obstetrical services. We find that the variance in physicians' weekend risk-adjusted c-section rates is 66 percent as large as the weekday variance, which implies that up to one-third of the weekday variation may be due to patient-physician matching based on preferences for a physician's treatment style. Physicians with high weekend c-section rates perform a relatively large number of scheduled weekday patient-choice (non high-risk) c-sections, which is consistent with women seeking out physicians who share their views regarding the net benefits of a c-section. Therefore, a considerable component of the variation could be welfare enhancing.

There is more than twice as much variation across physicians within a practice than one would expect if weekday patients were randomly assigned, which indicates that different members of a group practice are specializing in high- and low-risk patients. However, we cannot reject the hypothesis that performing more c-sections on "marginal" patients -- women who would receive c-sections if treated by physicians in their practices with favorable views of c-sections, but would not receive c-sections if physicians with less favorable views were on-call instead -- would have no effect on the mother's health outcome. This is consistent with physicians using c-sections appropriately.

[16] Without data on the babies' health outcomes, we cannot examine the combined health effect of a c-section for the mother and her baby.

References

American Board of Obstetrics and Gynecology (2005). http://www.abog.org/cert/cert.html.

Bikhchandani, Sushil, Amitabh Chandra, Dana P. Goldman, and Ivo Welch, 2001, "The Economics of Iatroepidemics and Quakeries: Physician Learning, Information Cascades, and Geographic Variation in Medical Practice," working paper.

Chandra, Amitabh, and Douglas Staiger, 2004, "Testing a Roy Model With Productivity Spillovers," NBER working paper 10811.

Chasin, Mark R., Robert H. Brook, R.E. Park, Joan Keesey, Arlene Fink, Jacqueline Kosecoff, Katherine Kahn, Nancy Merrick, and David H. Solomon, "Variations in the Use of Medical and Surgical Services By the Medicare Population," The New England Journal of Medicine 314(5): 285-290.

Epstein, Andrew and Sean Nicholson, 2005, "The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections," NBER working paper 11549.

Garicano, Luis, and Thomas N. Hubbard, 2003, "Specialization, Firms, and Markets: The Division of Labor Within and Between Firms," NBER working paper 9719.

Goyert, Gregory L., Sidney F. Bottoms, Marjorie C. Treadwell, and Paul C. Nehra, 1989, "The Physician Factor in Cesarean Birth Rates," *The New England Journal of Medicine* 320(11): 706-709.

Grant, Darren, 2004, "Information and Sorting in the Market for Obstetrical Services," *Health Economics*, forthcoming.

Gruber, J, J Kim, and D Mayzlin. 1999. "Physician fees and procedure intensity: The case of cesarean delivery." *Journal of Health Economics*. 18: 473-490.

Grytten, Jostein, and Rune Sorensen, 2002, "Practice Variation and Physician-Specific Effects," *Journal of Health Economics* 22(3): 403-418.

HealthGrades, 2004, *Hospital Report Cards™ Obstetric Services and Women's Health 2004 Methodology White Paper (2000- 2002 Data).* http://www.healthgrades.com/AboutUs/index.cfm?fuseaction=Ratings, accessed April 26, 2005.

Newhouse, Joseph P., 2002, "Why is There a Quality Chasm?" *Health Affairs* 21(4): 13-25.

Phelps, Charles E., and S. Parente, 1990, "Priority Setting for Medical Technology and Medical Practice Assessment," *Medical Care* 28(8): 703-723.

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

Phelps, Charles E., 2000, "Information Diffusion and Best Practice Adoption," in *Handbook of Health Economics*, vol. 1A, Joseph P. Newhouse and Anthony J. Cuyler, editors: 223-264. Amsterdam: Elsevier Science Publishers.

Roos, Noralou P., Gordon Flowerdew, Andre Wajda, and Robert B. Tate, 1986, "Variations in Physicians' Hospitalization Practices: A Population-Based Study in Manitoba, Canada," *American Journal of Public Health* 76(1): 45-51.

Stano, Miron, and Sherman Folland, 1988, "Variations in the Use of Physician Services by Medicare Beneficiaries," Health Care Financing Review 9(3): 51-58.

Welch, H. Gilbert, Mark E. Miller, and W. Pete Welch, 1994, "Physician Profiling: An Analysis of Inpatient Practice Patterns in Florida and Oregon," *The New England Journal of Medicine* 330(9): 607-612.

Wennberg, John E., and A. Gittelsohn, 1973, "Small Area Variations in Health Care Delivery," *Science* 182: 1102-1108.

Wennberg, John E., J.L. Freeman, and W.J. Culp, 1987, "Are Hospital Services Rationed in New Haven or Over-Utilised in Boston?," *Lancet* 1: 1185-1189.

Table 1: Sample Statistics for the Patient-level Cross-section Data Set (1999-2004)

	Weekend Laboring Deliveries (n = 145,303)		All Weekday Deliveries (n = 611,544)	
	Mean	Standard Deviation	Mean	Standard Deviation
Age	27.2	6.21	28.0	6.28
Black or hispanic	0.500	0.500	0.460	0.498
Medicaid or other public health insurance	0.501	0.500	0.460	0.498
Woman's health condition:				
Woman has had a previous c-section	0.043	0.203	0.150	0.357
Malpositioned fetus	0.029	0.168	0.071	0.257
Antepartum bleeding	0.009	0.094	0.017	0.130
Severe hypertension	0.005	0.072	0.013	0.114
Preterm gestation	0.067	0.220	0.075	0.263
Multiple gestation	0.006	0.078	0.014	0.118
Maternal soft tissue disorder	0.016	0.124	0.029	0.168
Macrosomia	0.016	0.127	0.030	0.170
Oligohydramnios	0.004	0.060	0.007	0.083
Polyhydramnios	0.015	0.123	0.039	0.193
Herpes	0.010	0.102	0.015	0.120
Uterine scar unrelated to prior c-section	0.001	0.021	0.002	0.047
Any of the above high-risk conditions	0.181	0.385	0.271	0.445
Mother went into labor	1.00	0.00	0.790	0.407
Delivery performed by c-section:				
- Scheduled, high-risk	0.00	0.00	0.106	0.307
- Scheduled, patient-choice	0.00	0.00	0.104	0.305
- Unscheduled (following labor)	0.096	0.294	0.088	0.284
Health outcome (continuous measure)	0.033	0.179	0.044	0.229
Indicator of bad health outcome	0.015	0.120	0.021	0.142

Table 2

OLS Coefficient Estimates of the Likelihood of Receiving a C-section in Florida, 1999-2004

	Weekend Laboring Deliveries		**All Weekday Deliveries**	
	Coefficient	Standard Error	Coefficient	Standard Error
Age	-0.0047**	0.0014	-0.0047**	0.0008
Age squared	0.0001**	0.00002	0.0001**	0.00001
Black or hispanic	0.0011	0.0025	0.0023	0.0015
Medicaid or other public health insurance	-0.016**	0.0025	-0.019**	0.0016
Woman's health condition:				
Woman has had a previous c-section	0.124**	0.0055	0.629**	0.0018
Malpositioned fetus	0.274**	0.0059	0.490**	0.0024
Antepartum bleeding	0.071**	0.011	0.313**	0.0047
Severe hypertension	0.180**	0.012	0.294**	0.0049
Preterm gestation	-0.057**	0.0042	-0.014**	0.0025
Multiple gestation	0.110**	0.015	0.165**	0.0057
Maternal soft tissue disorder	0.125**	0.0082	0.162**	0.0037
Macrosomia	0.198**	0.0074	0.277**	0.0033
Oligohydramnios	0.144**	0.017	0.145**	0.0074
Polyhydramnios	0.070**	0.0092	0.109**	0.0035
Herpes	0.0088	0.0090	0.178**	0.0045
Uterine scar unrelated to prior c-section	0.223**	0.047	0.454**	0.013
Observations	71,443		325,642	
R^2	0.25		0.61	

Notes: The dependent variable is an indicator for whether a woman received a c-section. We include a full set of physician indicator variables and omit the constant. Indicators are also included for whether a patient's race is missing, whether her age is missing, and various other health measures (uterine rupture, unengaged fetal head, congenital fetal CNS anomaly, cerebral hemorrhage, diabetes, chorioamnionitis, ruptured membrane > 24 hours, maternal hypotension, intrauterine growth restriction, maternal heart disease, asthma, maternal renal abnormality, and other maternal infection).

Table 3: Simulation Results

Weekend Laboring Deliveries, Among All Practices (n=189)

Patient Characteristics	Percentage of Groups Where Actual Variance is Significantly Larger Than Simulated Variance	Median of the Ratio of Actual-to-Simulated Variance
Predicted probability patients receive c-sections ($\hat{C}_j$)	22.8%	1.81
Proportion of patients with at least 1 of the 12 high-risk conditions	20.6%	1.65

Weekday Deliveries, Among Group Practices That Randomly Assign Weekend Deliveries (n=147)

Patient Characteristics	Percentage of Groups Where Actual Variance is Significantly Larger Than Simulated Variance	Median of the Actual to Simulated Variance Ratio
Predicted probability a patient will receive a c-section ($\hat{C}^d_j$)	36.7%	2.24
Proportion of patients with at least 1 of the 12 high-risk conditions	32.7%	2.20
Proportion of patients who are > 35 years of age	38.1%	3.19
Proportion of patients with a malpositioned fetus	20.4%	1.34
Proportion of patients who are a minority	54.4%	5.49
Delivery Method		
Proportion of deliveries that are:		
- scheduled c-sections	45.6%	3.26
- scheduled high-risk c-sections	32.7%	1.99
- scheduled patient-choice c-sections	41.5%	3.28

Notes: This table reports the results of a simulation of 1,000 iterations. In each iteration, a patient is randomly assigned to a physician in the patient's practice such that each physician receives his actual number of patients. After each iteration, we record the mean patient characteristics for each physician, and the variance across physicians within a practice in those characteristics. There are a total of 189 practices and 773 physicians that exceed our delivery threshold in the data set, which contains deliveries that occurred between 1999 and 2004. We test for significance at the 5-percent level using a one-tailed test. Specifically, if the actual within-practice variance exceeds the 95th percentile of the simulated variances for a particular practice, we conclude that the actual variance is significantly larger than the simulated variance.

Table 4

Panel A: Practice Characteristics Associated with Specialization

	Actual Variance Within Practice is Significantly Larger Than Simulated Variance?	
Weekend, All Practices (n=189)	NO	YES
Predicted probability a patient receives a c-section		
- number of physicians in the group	4.48	12.2**
- poor-paying patients as % of total	36.1	51.3**
Weekday, Practices That Randomly Assign Weekend Deliveries (n=147)		
Predicted probability a patient receives a c-section		
- number of physicians in the group	3.49	5.79**
- poor paying patients as % of total	37.8	33.9
Proportion of patients who are > 35 years of age		
- number of physicians in the group	3.27	5.85**
- poor-paying patients as % of total	37.4	34.6
Proportion of patients who receive a scheduled patient-choice c-section		
- number of physicians in the group	3.94	5.09**
- poor payers as % of total patients	38.3	33.6

Panel B: Correlations Between Measures of Weekday Specialization for Practices That Randomly Assign Weekend Deliveries

Specialization is measured as the ratio of the actual variance across physicians within a practice to the median of the 1,000 simulated variances across physicians in that practice.

	A	B	C	D	E	F	G
Patient Characteristics							
A: Patients' c-section propensity	1.00						
B: High-risk patients	0.37	1.00					
C: Malpositioned fetus	0.62	0.26	1.00				
D: > 35 years of age	0.23	0.18	0.23	1.00			
E: Minority patients	0.01	0.09	0.00	0.56	1.00		
Delivery method							
F: Scheduled high-risk c-section	0.65	0.38	0.62	0.15	0.04	1.00	
G: Scheduled patient-choice c-section	0.40	0.09	0.20	0.10	0.08	0.27	1.00

Notes: ** = significantly different from the mean value for the practices that do not specialize, at the 5-percent significance level.

Table 5

Sorting on Unobservables: Variation Between Physicians in Their Treatment Styles

Standard Deviation of the Difference Between a Physician's Risk-adjusted C-section Rate and the Market Average, Among Physicians in Group Practices

	All MDs (n = 773)	MDs in Practices That Randomize Weekend Deliveries (n = 481)
A: All weekday deliveries	0.0425	0.0432
B: Weekend laboring deliveries	0.0337	0.0350
F-test that A > B	0.000	0.000
Ratio of weekend to weekday variance	0.63	0.66

Notes: Both variables are measured as differences from the local health district, or regional, mean. The data are for deliveries performed between 1999 and 2004.

Table 6

Panel A: Correlation Coefficients to Test Mechanism of Sorting on Observed Patient Characteristics

	Predicted c-section probability for MD j's weekday patients – his practice's average
All physicians (n = 773)	
MD j's – his practice's average risk-adjusted weekend c-section rate	0.022 [0.547]
MDs in Practices that randomize weekend deliveries (n=481)	
MD j's – his practice's average risk-adjusted weekend c-section rate	0.030 [0.509]

Panel B: Correlation Coefficients to Test Mechanism of Sorting on Unobserved Patient Characteristics

	MD j's – his practice's avg. quantity of weekday scheduled c-sections		
All physicians (n=773)			
	High-risk	Patient-choice	Total
MD j's – his practice's average risk-adjusted weekend c-section rate	0.023 [0.517]	0.160** [0.000]	0.088** [0.014]
MDs in practices that randomize weekend deliveries (n=481)			
	High-risk	Patient-choice	Total
MD j's – his practice's average risk-adjusted weekend c-section rate	0.038 [0.411]	0.094** [0.038]	0.071 [0.120]

Notes: p-values are in brackets. ** = significantly different from zero at the 5-percent level.

Table 7

IV Estimates for the Effect of a C-section on the Mother's Health for Weekend Laboring Deliveries

	All Patients	Patients in Practices That Randomize Weekend Deliveries
OLS coefficient on c-section indicator	0.084**	0.056**
	(0.0017)	(0.0020)
Instrumental Variables Estimate		
First-stage estimate of OBdiff	0.266**	0.079*
	(0.031)	(0.043)
Second stage:		
C-section indicator	0.112	-0.197
	(0.100)	(0.394)
Age	-0.0016**	-0.0032
	(0.0007)	(0.0022)
Age squared	0.00002*	0.00006
	(0.00001)	(0.00004)
Minority	0.0069**	0.0059**
	(0.0014)	(0.0016)
Poor payer	0.0007	-0.0062
	(0.0020)	(0.0059)
# of patient observations	145,503	84,473
# of physicians	773	481

Notes: The dependent variable in the OLS regression is a mother's health outcome, where a larger value indicates a worse outcome. The dependent variable in the first stage of the two-stage least squares model is an indicator variable for whether a woman received a c-section, and the dependent variable in the second stage is a mother's health outcome. A comprehensive set of indicator variables are included for a woman's health at the time of the delivery and the status of the pregnancy in both the first- and second-stage regressions. Practice and year fixed effects are also included as controls. OBdiff is the difference between the actual proportion of a physician's weekend laboring patients who received a c-section and the predicted proportion based on the patients' observed characteristics, with the effect of patient i removed from the calculation.

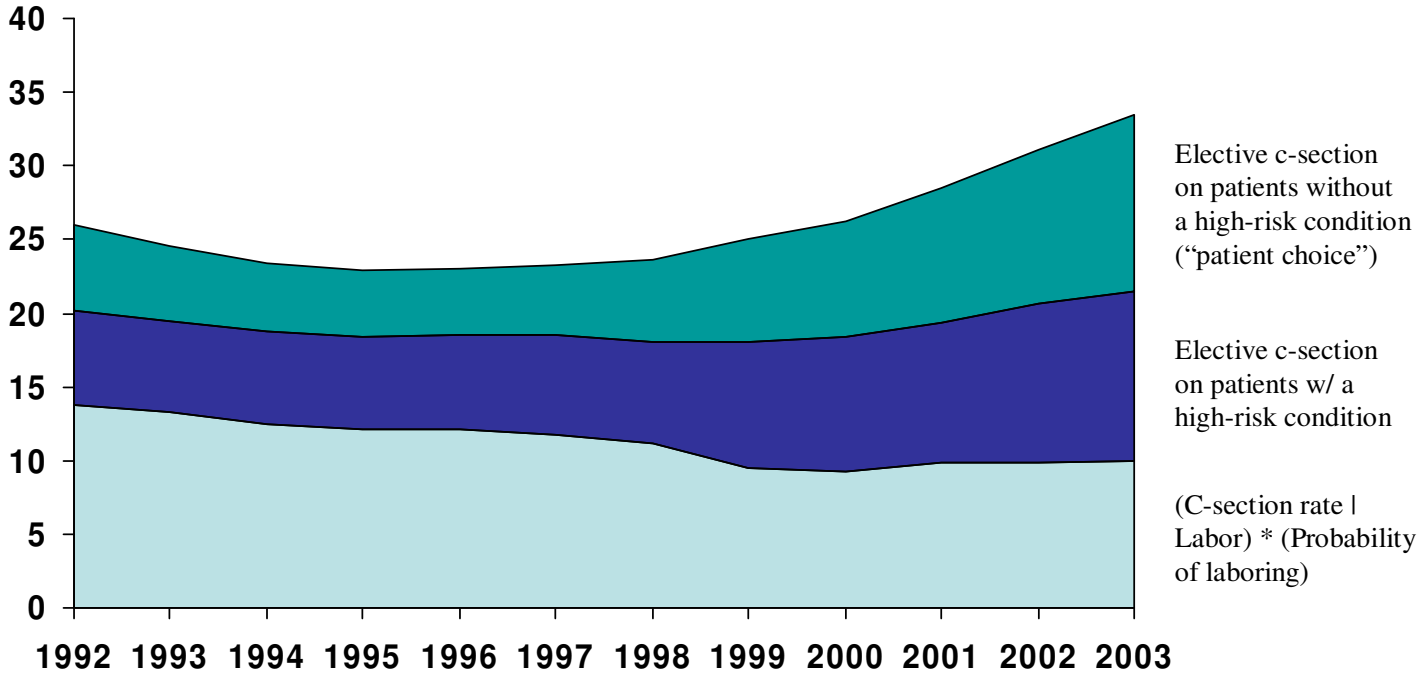
Appendix: Creating a Continuous Health Outcome Variable for the Mother's Health

Coefficient Estimates of a Woman's Length of Stay (in Days) in the Hospital in Florida

	Vaginal Delivery		**Cesarean Section**	
Adverse Outcome	Coefficient	S.E.	Coefficient	S.E.
Respiratory failure	4.856	0.299	3.592	0.233
Acute renal failure			3.197	0.429
Acute edema of lung	2.808	0.183	2.736	0.200
Postoperative infection			2.471	0.280
Major puerperal infection	1.752	0.060	1.745	0.046
3rd stage post-partum hemorrhage	0.513	0.044	1.486	0.163
Gastrointestinal complications			1.393	0.140
Respiratory complications			1.152	0.207
Other complications	1.621	0.119	1.140	0.183
Complication of OB surgical wound	0.839	0.085	1.060	0.068
Cardiac complications			1.042	0.385
Other specified complications			1.039	0.163
Other OB wound surgical complication			1.026	0.082
Disruption of c-section wound			0.893	0.117
Delayed post-partum hemorrhaging	0.663	0.042	0.872	0.272
Coagulation defects	0.682	0.065	0.835	0.103
Urinary complications	0.419	0.194	0.723	0.285
Secondary thrombocytopenia	0.489	0.147	0.626	0.230
Anesthesia complication	0.989	0.121	0.613	0.173
Puncture or laceration by an instrument	0.767	0.237	0.581	0.168
Hemorrhage or hematoma	0.945	0.247	0.340	0.197
Post-partum hemorrhage w/in 24 hours	0.331	0.017	0.313	0.061
Acute anemia due to loss of blood			0.299	0.035
Retained placenta			0.219	0.184
Anemia following the delivery			0.111	0.034
3rd degree perineal laceration	0.153	0.013		
4th degree perineal laceration	0.228	0.019		
Laceration of the cervix	0.434	0.048		
High vaginal laceration	0.131	0.019		
Injury to pelvic organs	-0.031	0.015		
Constant	2.487	0.017	3.720	0.060
Observations	142378		44030	
R-squared	0.088		0.101	

Figure 1

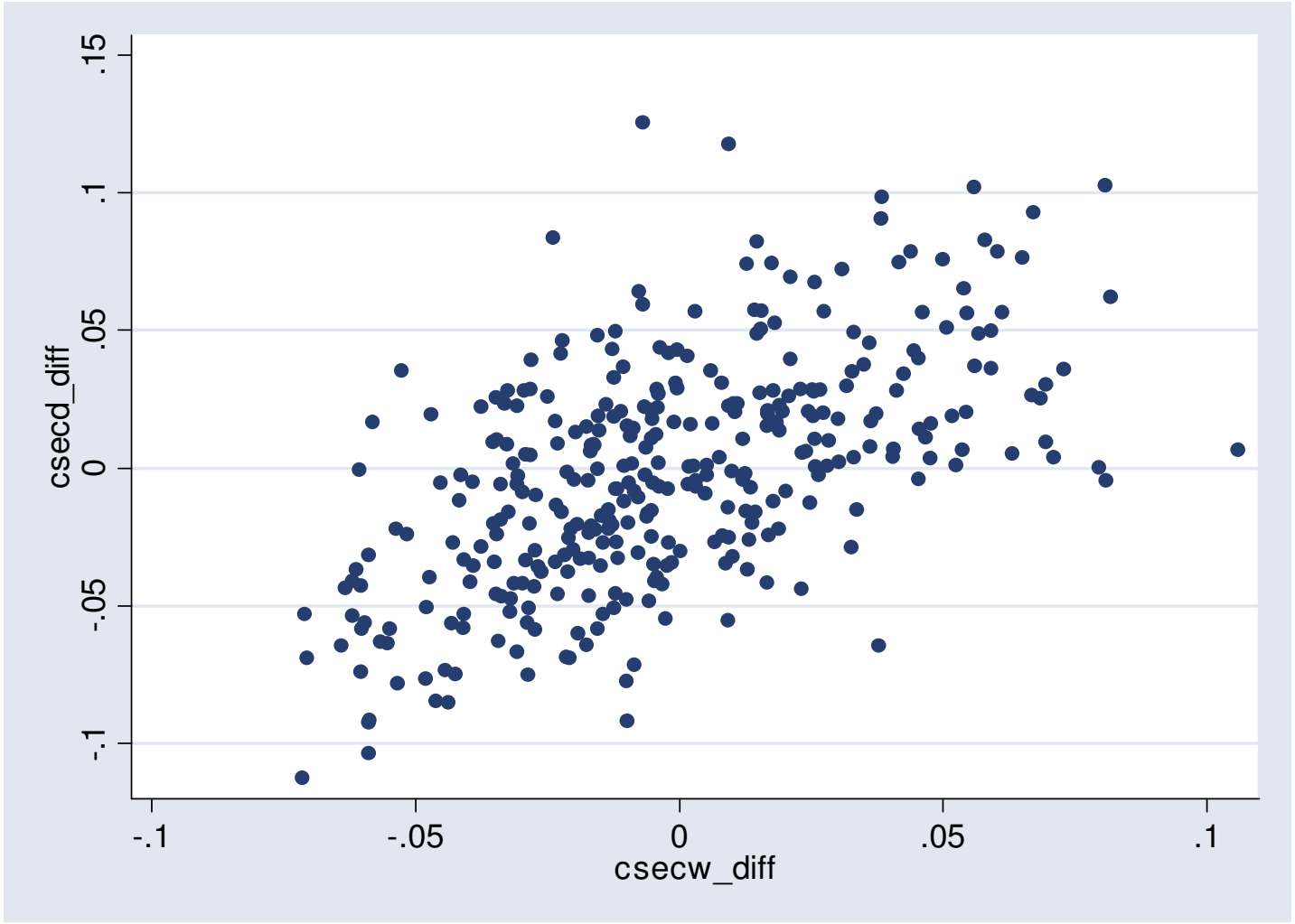
Percentage of C-section Deliveries in Florida by Type, 1992-2003



Note: a delivery where a woman does not labor is classified as an elective c-section. If a woman has any of the first 12 health conditions listed in Table 1, we consider her to be a "high-risk" patient.

Figure 2

Relationship Between a Physician's Weekend and Weekday Risk-adjusted C-section Rates



Difference between a physician's and the market average risk-adjusted WEEKEND c-section rate