NBER WORKING PAPER SERIES

PROFESSIONAL PARTNERSHIPS AND MATCHING IN OBSTETRICS

Andrew Epstein
Jonathan D. Ketcham
Sean Nicholson

Working Paper 14070
http://www.nber.org/papers/w14070

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2008

We appreciate helpful comments from David Bradford, David Meltzer, Doug Staiger, Jonathan Skinner, and seminar participants at NBER, Syracuse University, the 2005 Annual Health Economics Conference, the 2006 American Society of Health Economists conference, and the 2007 American Economic Association annual meeting. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Professional Partnerships and Matching in Obstetrics
Andrew Epstein, Jonathan D. Ketcham, and Sean Nicholson
NBER Working Paper No. 14070
June 2008, Revised September 2008
JEL No. D83,I12,J44,L15,L25

ABSTRACT

Theory indicates that internally-differentiated professional partnerships can promote matching between heterogeneous consumers and professionals, particularly when consumers have imperfect information or markets have barriers to referrals between firms. We test this in obstetrics markets, relying on random assignment of patients to physicians to generate unbiased measures of a physician's treatment style and skill, and on simulations to measure a physician's specialization. Consumers match to professionals along all three dimensions -- specialization, style and skill -- based on consumers' observed characteristics and unobserved preferences. We conclude that internally-differentiated partnerships promote matching in ways that improve consumers' welfare and health.

Andrew Epstein
School of Public Health
Yale University
60 College St, Room 301
New Haven, CT 06520
andrew.epstein@yale.edu

Jonathan D. Ketcham
Assistant Professor
School of Health Management and Policy
W.P. Carey School of Business
Arizona State University
300 E. Lemon Street
Tempe, AZ 85287-4506
ketcham@asu.edu

Sean Nicholson
Assistant Professor
Department of Policy Analysis and Management
Cornell University
123 Martha Van Rensselaer Hall
Ithaca, NY 14853
and NBER
sn243@cornell.edu

I. Obstetrics Markets

Many theoretical models of professional service partnerships have been developed to understand better the formation of such firms, their boundaries and their scopes. These models have focused on the benefits to the professionals or firms themselves. For example, partnerships' profit-sharing can effectively spread the risk of demand shocks across the members (Gaynor and Gertler 1995). Alternatively, Chou (2007) proposes that human-capital intensive firms merge to establish a single brand identity that obfuscates the individual market value of the individuals or firms within the larger firm. The rationale for mergers between firms in this model is the merged firm's ability to capture more of the investment in human capital such as knowledge, skill, or specialization. Other models have considered the ability of partnerships to signal higher quality to consumers in markets where quality is not readily observable, which they achieve by screening out low-quality professionals through their hiring decisions (Levin and Tadelis 2002) or by creating incentives for members to invest in quality-improving human capital (Morrison and Wilhelm 2004).

Other theoretical models have focused on the advantages of firms with heterogeneous workers. Firms with specialized workers can coordinate their production (Alchian and Demsetz 1972; Garicano 2000; Hart and Moore 2005), facilitate the flow of information among specialists (Bolton and Dewatripont 1994), and match them with the most appropriate tasks (MacDonald 1982). In professional services markets, "tasks" can refer to individual, heterogeneous customers. The welfare implications of product differentiation in markets with heterogeneous consumers depend on the ability of consumers to match with products. In markets with imperfect or asymmetric information, firms can facilitate matching of consumers with goods and services by reducing consumers' search and diagnosis costs (Wolinsky 1993) or by eliminating or reducing disincentives to refer consumers (Garicano and Santos 2004). In Garicano and Santos (2004), under many but not all conditions, professional partnerships create incentives for within-firm referrals that enhance matching between specialists and uninformed consumers beyond those achieved through (post-diagnosis) spot markets or (pre-diagnosis) retainer contracts between professionals in different firms.

In this paper we empirically consider the consumer benefits from firms of professionals who are differentiated in specialization, style and skill. We study these effects conditional on the existing makeup of firms rather than testing the various models that explain the organization of professionals into firms. Specifically we analyze whether obstetrics partnerships promote matching between heterogeneous consumers (patients) and professionals, as proposed by Garicano and Santos. In this context, “partnerships” refer to physician group practices that share costs and revenues in ways that permit referrals without violating anti-kickback regulations that prevent payments between firms. The three specific types of matching we consider in this paper are: 1) whether a patient with a specific health condition is assigned to a physician who specializes in treating that condition; 2) whether a patient who prefers to be treated intensively is assigned to a physician with an intensive treatment style or skill advantage for that treatment; and 3) whether a patient who is medically appropriate for cesarean section is assigned to a physician who is skilled at performing that treatment.

One empirical challenge to identifying producer differentiation, subsequent matching and their welfare implications is the endogenous matching of workers with jobs (Ackerberg and Botticini 2003). We rely on simulations to quantify the extent of specialization, and we capitalize on the random pairing of physicians and patients on weekends to measure physicians’ styles (i.e. propensity to perform a cesarean section) and skills (as revealed by maternal health outcomes) without bias due to matching on unobserved patient or physician characteristics. To do this, we exploit the fact that because the timing of childbirth is uncertain and physicians’ weekend call schedules are predetermined, most women being treated by a physician in a group practice who go into labor on the weekend are randomly assigned to a member of the practice who is on call at the time. We first confirm this feature and then use it to determine whether group practices match physicians skilled at performing cesarean sections with patients clinically appropriate to receive the procedure, as well as whether patients with preferences for being treated by cesarean section match with physicians who have the appropriate treatment style or skill advantage. We also rely on the random pairing to identify the health benefits of matching specialization by comparing the health outcomes of high-risk patients randomly assigned to physicians who specialize in that particular health condition to the outcomes of other high-risk patients in the practice who were randomly assigned to non-specialists.

We find that the proportion of high-risk patients matched with a specialist is larger among internally-differentiated group practices (i.e., practices where some but not all physicians specialize) than among solo practitioners or homogeneous groups for 11 of the 12 specific health conditions we examine. This additional matching is most pronounced for health conditions that develop during a pregnancy, when physicians' financial disincentives to refer across firms are largest. Although in most cases this matching does not affect maternal health outcomes during the delivery or the subsequent hospital stay[1], when it does affect maternal health, the magnitudes of the estimated effects are large.

We also find that physicians in group practices vary from their fellow group practice members in treatment style and skill levels. Although we do not find clear evidence that groups match patients with preferences for cesarean section with physicians who have a high propensity to perform them (style), our results indicate that they are treated by physicians in the group who are relatively more skilled at cesarean sections. Likewise, group practices direct patients for whom cesarean sections are more clinically appropriate to these more skilled physicians. Taken together, our results indicate that differentiation among physicians within an obstetrics practice promotes patient-physician matching in ways that improve welfare.

This paper is structured as follows. Section II provides an overview of obstetrics markets and Section III describes the data. In Section IV we examine whether group practices promote

[1] We are not able to examine the effect of specialization on babies' health outcomes due to data limitations.

physician specialization and whether differentiated groups enhance physician-patient matching based on patients' health conditions relative to patients treated by obstetricians in homogeneous practices. We then analyze the effects of matching and mismatching on mothers' delivery outcomes in Section V. In Section VI we examine whether patients in group practices match with physicians based on patient treatment preferences and physicians' treatment styles, as well as based on physicians' skills and patients' clinical characteristics. Section VII concludes.

II. Obstetrics Markets

Obstetrics markets are characterized by heterogeneous consumers, asymmetric information, heterogeneous professionals often organized into partnerships, and disincentives for referrals between practices. These markets differ from the assumptions of the Garicano and Santos (2004) model in several important ways. First, "self-referral" regulations commonly known as the "Stark law"[2] prohibit both spot markets and retainer contracts between physicians who are not members of the same group practice. At the same time, payment for prenatal care typically is bundled with payment for the delivery, with all or almost all of the payments made to the physician performing the delivery.[3] One implication of this is that a physician's expected profit-margins are lower at the outset of the pregnancy than they are closer to the delivery. Because partnerships typically share revenues, members of a group medical practice have no or smaller financial disincentives to avoid referrals between members of the same group. Together these provide strong disincentives for a physician to refer outside of but not within his own practice, particularly later in the pregnancy when the marginal expected profits are higher. As a result, groups might be especially effective at promoting matching on clinical characteristics that become known later in a pregnancy, such as preterm gestation, malpositioned fetus, pregnancy-

[2] The most recent regulations are officially titled, "Physicians' referrals to health care entities with which they have financial relationships," and are available at http://frwebgate.access.gpo.gov/cgi-bin/getpage.cgi?dbname=2001_register&position=all&page=856.
[3] For one example, see http://www.tuftshealthplan.com/providers/pdf/payment_policies/obgyn.pdf. Accessed April 25, 2008.

related hypertension, congenital anomalies or chromosomal abnormalities.[4] In addition to these physician disincentives for late-term referrals across groups, high search costs and switching costs might discourage patients from changing practices (Wolinsky 1993), which creates an additional option value from selecting a partnership over a solo physician. Moreover, while Garicano and Santos (2004) assume consumers are uninformed, expecting mothers are likely to acquire information regarding their health condition and the characteristics of obstetricians through a number of sources, including referring primary care physicians, their own experiences, word-of-mouth, or externally-produced comparisons such as quality report cards (Hoerger and Howard 1995). Thus we would expect some matching to occur between patients and obstetricians in homogeneous practices (including solo physicians) based on information that is available to patients and their referring physicians. Although we cannot observe referrals directly, if the same amount of information exists *ex ante* in the market regarding physicians in homogeneous groups and those in groups with differentiated physicians, the additional matching that occurs in differentiated groups would be due purely to within-group referrals.

Groups with differentiated physicians have the potential to promote matching, but relative to homogeneous groups they also could foster more mismatches because of how group practices schedule obstetricians, hospital staffing patterns, and the emergent nature of childbirth. 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, although some physicians make themselves available to perform their patients' deliveries during the week (but not weekend) even when not on call (Schauberger, Gribble and Rooney 2007).[5] Thus many patients who go into labor naturally (i.e., without being induced) are randomly assigned to a member of the group practice, particularly on weekends.

---

[4] Although risk factors observed at the outset might provide evidence to physicians about the likelihood of developing hypertension from pregnancy, to our knowledge women have equal observable risk of having a malpositioned fetus.

[5] In other practices, on weekdays physicians perform deliveries for patients who begin labor between 7am and 5pm, but rely on physicians on "night call" to cover deliveries during the remaining hours.

If a patient or the physicians in the group practice want to guarantee that a particular physician performs a patient's delivery, they can either schedule a cesarean section or schedule labor to be induced when the preferred physician is on-call at the hospital. Induced deliveries and scheduled cesarean sections are much more likely to occur on a weekday than a weekend because hospitals are staffed more heavily during the week, and many physicians value leisure time on the weekend more highly. Scheduled cesarean sections and induced deliveries are common, which means there are many opportunities for intentional matching of patients and physicians on weekdays. During our 1999-2004 sample period, 17.1 percent of all deliveries in Florida and New York were scheduled cesarean sections, and almost all of these occurred on a weekday. Nationally, 20 percent of deliveries in the United States in 2000 were induced, and these were 2.3 times more likely to occur on a weekday than a weekend (Martin et al. 2003).

III. Data

We construct our sample from the Florida and New York hospital discharge data sets for the 1999 to 2004 time period. Our analytic data set contains information on 1.6 million deliveries that occurred at every non-federal, short-term acute care hospital in those states.[6] We observe the mother's demographic information (age, race), her insurance coverage (e.g., HMO, Medicaid), codes for the principal and secondary diagnoses, procedure codes that allow us to determine whether the baby was delivered vaginally or via cesarean section, a unique physician identifier that is consistent across hospitals and years, a unique and consistent hospital identifier, and the quarter and year the patient was discharged. The diagnoses codes allow us to control for objective health conditions that affect the probability a physician will perform a cesarean section (e.g., whether a woman had a cesarean section prior to this delivery, whether the fetus was malpositioned during the delivery, and whether the labor occurred before the fetus was full-term).

---

[6] There were a total of 2.38 million deliveries in Florida and New York hospitals between 1999 and 2004. About 740,000 of these deliveries are excluded from the analysis because the physician performed fewer than 100 weekend laboring deliveries, a woman received a scheduled cesarean section on a weekend, or we could not identify the physician who performed the delivery in the physician databases.

Women are much more likely to receive a cesarean section in these situations (Gregory et al. 2002).[7]

To minimize concerns about measuring physician specialization, treatment styles and skill with error, throughout the paper we restrict the analysis to physicians who performed at least 100 deliveries between 1999 and 2004 for patients who went into labor and delivered on weekends. We rely on the work of others (Henry et al. 1995; Gregory et al. 2002) to determine whether a woman went into labor. Women who delivered vaginally or had codes indicating fetal distress, labor abnormalities, cord prolapse, or a breech converted to vertex presentation were interpreted as having labored.

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 telephone number(s) and office name in April 2005 from the Florida State Licensure Database and American Medical Information, a private company that tracks physicians for general marketing purposes. Where telephone numbers or office names were missing, we relied on internet searches conducted in September 2005 and January 2007 to identify them. If a physician had a unique telephone number in our sample, we concluded that he is a solo practitioner; physicians who shared the same telephone number were considered to be in the same group practice.[8] Identical practice names and physicians with multiple telephone numbers were used to identify two or more different locations of the same practice.[9]

A total of 1,655 physicians (930 in New York and 725 in Florida) performed 100 or more weekend laboring deliveries between 1999 and 2004. On average, each physician performed 796

[7] Regression results below confirm this.

[8] Because physicians who performed fewer than 100 weekend deliveries between 1999 and 2004 are omitted from the sample, some of the physicians who qualify as being solo practitioners may in fact be in a group practice with other low-volume obstetricians, or with physicians who do not deliver babies (e.g., gynecologists.)

[9] We measure physicians' practice affiliations with error if they changed between the start of our sample period (1999) and when we identified them (2005 and 2007).

weekday and 187 weekend laboring deliveries over the six-year period. We identify 12 high-risk health conditions (e.g., twins) that substantially increase the likelihood that a pregnant woman or her baby will experience an adverse health outcome during delivery (Gregory et al. 2002). Table 1 reports sample means and standard deviations for patients treated by physicians who met the threshold volume of weekend deliveries, separately for weekend and weekday deliveries. Overall, 26.2 percent and 17.3 percent of women who delivered during the week or on the weekend, respectively, had at least one of these 12 conditions. The presence of a high-risk condition substantially increases the probability that a woman will receive a cesarean section, which is a surgical procedure, rather than delivering vaginally, and increases the probability that the mother or the baby will have a negative health outcome during or shortly after the delivery. For example, among women who delivered on a weekday, 54.6 percent of women who had a high-risk condition received a cesarean section, versus 21.6 percent of women without a high-risk condition.

IV. Matching on Specialization

Obstetricians can specialize in treating particular types of patients. For example, some obstetricians perform a relatively large number of deliveries for women who have specific health risks such as hypertension (high blood pressure, or pre-eclampsia), a malpositioned fetus (e.g., a breech baby), or antepartum bleeding, while other obstetricians avoid such cases. A physician who specializes in treating a specific health condition might produce better health outcomes for the mother or the baby than a physician who avoids such cases, either due to innate skill differences between obstetricians or from learning-by-doing (e.g., technical proficiency acquired through experience) (Britt, Eden and Evans 2006). If this occurs, then the welfare effects of specialization depend on the extent to which women with a particular health condition are able to match with a physician who specializes in that health condition.

We compare the number of each type of high-risk delivery[10] each obstetrician in Florida and New York actually performed between 1999 and 2004 on weekdays, when patients have opportunities to choose the delivering physician by scheduling a cesarean section or requesting an induced delivery, to the number of deliveries expected if patients were randomly assigned to an obstetrician in the market. An obstetrician is then categorized as a "specialist" if he performs significantly more high-risk deliveries than expected, an "avoider" if he performs significantly fewer high-risk deliveries than expected, or a "generalist" otherwise. This allows us to compare the extent of specialization in group practices versus solo practices. It also defines whether the practice consists of differentiated or homogeneous physicians. We use this to consider the effects of within-practice differentiation on the proportion of patients who match (i.e., high-risk patients assigned to specialist physicians or low-risk patients assigned to avoiders) or mismatch (i.e., high-risk patients assigned to avoiders or low-risk patients assigned to specialists).

In each of 1,000 iterations of a simulation, each weekday patient is randomly assigned to a physician in the patient's Local Health Council District (LHCD) in Florida, or Health Service Area (HSA) in New York,[11] such that each physician receives his actual number of weekday patients for each year in the sample period. A physician is defined as a specialist in each of the 12 high-risk conditions if the prevalence of that condition among his actual patients exceeds the 99th percentile of the simulated prevalence of that condition; an "avoider" if his actual prevalence falls below the 1st percentile in the simulation; and a generalist otherwise.[12]

In Table 2 we report the proportion of physicians who are specialists and avoiders for each of the 12 high-risk conditions, as well as for the summary indicator of whether a woman has

[10] A delivery is counted separately for each high-risk condition a patient has.
[11] The 11 LHCDs in Florida were formed by the state for purposes of health planning (see http://www.flhealthplanning.org/). The eight HSAs in New York were formed for a similar purpose. Each consists of one or more contiguous counties.
[12] Our results are qualitatively similar if we instead set the thresholds at the 95th and 5th percentile, or if we define the market as the hospital where the physician performs the majority of his deliveries. The LHCD/HSA and the hospital approaches represent extremes of the possible market, because the former definition is likely to be too large and the latter too small.

any of the 12 high-risk conditions. The conditions are arranged in descending order according to the mean number of deliveries performed by a specialist physician over the six-year period (shown in the last column). The results in the first column of Table 2 indicate that specialization is infrequent. Based on the unweighted average across the 12 conditions, 13 percent of physicians are specialists and 22 percent are avoiders. There is variation across conditions, from 7 percent of physicians specializing in pregnancies with congenital anomalies or chromosomal abnormalities up to 20 percent specializing in patients with any of the 12 high-risk conditions. The percentage of physicians who are avoiders likewise ranges from a low of 11 percent for antepartum bleeding to a high of 32 percent for uterine scarring. The difference in the volume of deliveries for patients with each high-risk condition between a specialist and an avoider is substantial, as indicated in the last column. Specialists perform two to 127 times more deliveries, on average, relative to avoiders.

In the second and third columns of Table 2 we report the proportions of physicians who are specialists and avoiders, separately for solo practitioners and members of a group practice. Physicians in group practices are slightly more likely than solo practitioners to specialize and are less likely to be avoiders, although we cannot determine whether groups purposefully recruit physicians different from their existing members or whether the organizational structure facilitates differentiation. A larger proportion of physicians in group practices specialize in patients who have a multiple birth, congenital anomalies, or hypertension, predominantly from pregnancy-related severe hypertension, relative to solo practitioners. None of these risk factors is observable until later in a pregnancy,[13] which is consistent with groups' referring relatively more patients within the group when it is particularly costly to refer to a different practice. A smaller proportion of physicians in group practices are avoiders for 8 of the 13 conditions relative to solo

[13] Of the 12 risk factors for which we have data, malpositioned fetus is the only additional one that is not observable until later in the pregnancy.

practitioners, whereas the opposite is true for only a single condition. Thus the random assignment of patients to on-call physicians appears to limit the number of avoiders in groups.[14]

With these measures of physician specialization, for each risk factor we then determine whether group practices are homogeneous (i.e., all specialists, all generalists, or all avoiders) or differentiated (e.g., a specialist and two generalists). In column 6 and column 7 of Table 2 we report the proportion of physicians who are specialists and avoiders separately for homogenous groups and differentiated groups. Across all health conditions, the proportion of physicians in differentiated groups who are specialists (17 percent) is more than twice that of homogenous groups (7 percent), and the proportion of physicians in differentiated groups who are avoiders is almost twice as large (25 percent versus 13 percent).

To test whether differentiated groups facilitate matching and minimize mismatching, in Table 3 we report the proportions of high-risk patients delivering on weekdays who are matched with specialists and mismatched with avoiders, separately for patients of differentiated group practices versus patients of solo practitioners or homogeneous groups. This indicates how many more high-risk patients are assigned to an appropriate physician in practices that have chosen to differentiate themselves and can refer between partners versus practices that by choice cannot refer within the practice (solo practitioners) or have chosen not to differentiate from one another (homogeneous group practices). On weekdays, differentiated groups achieve much greater matching than solo practitioners and homogeneous groups. The proportion of high-risk patients who are assigned to a specialist is higher among patients of differentiated group practices overall and for 11 of the 12 individual high-risk health conditions. On average, the share of high-risk patients who match with a specialist is 12 percentage points higher in differentiated group versus solo or homogeneous group practices. Four of the five largest differences in matching rates occur

[14] About 25% of deliveries performed by physicians in solo practices were for mothers with at least one of the 12 high-risk conditions, compared with 26% for physicians in group practice, suggesting that the differences in specialization do not stem from solo practitioners' positioning themselves to have healthier patients in general.

in the conditions that develop late in a pregnancy, where members of a group practice do not face the financial disincentives to refer that physicians in homogeneous practices do: multiple fetuses (28 percentage point difference), malpositioned fetus (23 percentage point difference), pregnancy-related severe hypertension (17 percentage point difference), and preterm gestation (15 percentage point difference).

High-risk patients in differentiated practices are slightly less likely to be assigned to avoiders (i.e., to mismatch) than those in homogeneous practices. Patients of differentiated group practices have a five percentage point lower mismatch rate than patients in homogeneous practices for the summary high-risk health measure, and the differentiated group mismatch rate is significantly lower for 5 of the individual measures of patient risk. These results are consistent with models of partnerships that form to lower the barriers to referrals among differentiated members (Garicano and Santos 2004).[15]

V. Matching and Mothers' Health Outcomes

Given these differences in matching and mismatching due to within-firm specialization, we next analyze whether and how matching and mismatching affect mothers' health during the delivery and subsequent hospital stay. Our identification of the health effect relies on random pairings of particular types of patients (based on whether they have a specific high-risk health condition) with a given type of physician (specialist, generalist, or avoider for that health condition) to eliminate differences in unobservable patient characteristics between physicians in a group due to referrals or other sources of patient-practice matching. As a result, we restrict the sample to group practice patients who delivered on the weekend, because typically these patients are randomly assigned to a physician within the group.

[15] We also compared the matching rates separately for patients who are uninsured or have Medicaid insurance versus patients with private health insurance, and found them to be similar. This suggests that maternal income and education do not appear to be important for creating matches, and that the benefits of within-group referrals do not vary by socioeconomic status. These results are available from the authors by request.

A small number of weekend deliveries are scheduled for a cesarean section or inducement, where scheduling can permit the patient and the physician to match intentionally.[16] Thus we further restrict the weekend sample to deliveries where a woman goes into labor, thereby excluding scheduled weekend cesarean sections and reducing the amount of non-random pairing in this sample. Because we cannot observe from the discharge data whether a delivery was induced, a small part of the weekend laboring sample might contain non-randomly-paired patients and physicians. To ameliorate the potential for bias, we further restrict the sample to groups that simulations indicate assign their weekend laboring patients randomly. The analysis considers only within-group differences, because call schedules assign patients to physicians randomly only within a group practice, and patient sorting into groups is likely to occur on the basis of unobserved characteristics that affect outcomes. As a result, only internally-differentiated groups can be used to identify the health benefits of matching.

Whether a group is internally-differentiated and randomly assigns its patients is determined separately for each health condition we examine. To identify and exclude any groups that do not randomly pair weekend patients with physicians, we perform 1,000 iterations of a simulation that randomly assigns a group's weekend, laboring patients to the physicians in that group, where the yearly number of simulated patients assigned to each physician equals his actual volume of weekend laboring cases. After each iteration we record the mean patient characteristics for each physician (e.g., the proportion of his patients who have antepartum bleeding, severe pregnancy-induced hypertension, or a malpositioned fetus). Group practices with at least one physician whose actual mean for a particular health condition is below the 5th percentile or above the 95th percentile of the physician-specific distribution of the simulated mean of that health condition are considered to violate the assumption that weekend patients are randomly assigned for that health condition, and are excluded from the outcomes analysis. We perform this test for

[16] Previous studies indicate that most induced deliveries nationally occur on weekdays, when hospitals are staffed more heavily (Martin et al. 2003).

ten of the 12 high-risk health conditions separately as well as for an indicator of whether a patient has any of the 12 conditions. We omit the two health conditions where specialists are most infrequent and have very low patient volumes: uterine scarring (a mean of 7.5 deliveries over a six-year period for specialists) and congenital anomalies or chromosomal abnormalities (a mean of 6 deliveries over a six-year period for specialists).

To measure mothers' outcomes, we use a coding system established by HealthGrades (2007) to determine whether or not a woman experienced an adverse outcome during the delivery or subsequent hospitalization. This system identifies 66 discrete diagnosis codes for adverse outcomes for a cesarean section and 64 discrete diagnosis codes for adverse outcomes for a vaginal delivery. Examples include injury to pelvic organs, rupture of the uterus, 4th degree perineal laceration, anemia due to acute loss of blood, or major puerperal infection. In our sample, adverse outcomes for the mother during or shortly after delivery occurred for 10.3 percent of deliveries where a woman went into labor and delivered on the weekend. To create a single summary measure of mothers' outcomes that allows the utility loss to vary across the different adverse events, we conduct regressions of the form

(1) $O_i = \phi_0 + \phi_1 R_i + F_i + \varepsilon_1$

where O_i is the number of days that a woman spends in the hospital following a delivery, R_i is a vector of indicator variables for whether the woman experienced each of the adverse events specified by Health Grades, and F_i represents hospital fixed effects. This approach assumes that the utility loss from an adverse event is proportional to the marginal effect of that event on the length of a mother's hospital stay, controlling for the hospital of admission.

We estimate these regressions separately by state and delivery mode (i.e., vaginal and cesarean section) using data from 1998, the year prior to the beginning of our sample. Results are reported in Appendix A. In the cesarean section regressions, the coefficients are greater than 0.50 and significantly different from zero for 20 of the adverse outcomes in Florida and 19 in New

York, whereas in both Florida and New York 11 events in the vaginal delivery regressions have a positive coefficient of that magnitude. That is, these events are associated with a woman spending at least an extra half day in the hospital following the delivery. To generate a single, continuous measure of each woman's health outcome, we multiply the vector of adverse outcome coefficients from the 1998 regressions by the vector of indicator variables for whether the woman in the 1999-2004 sample actually experienced each adverse event, so that larger values indicating a worse outcome.[17]

With measures of patient types, physician types, and mothers' outcomes, we analyze the impact of matching on mothers' health with regressions of the form

$$(2)\quad \hat{O}_i = \sigma_0 + \sigma_1 H_i + \sigma_2 S_j + \sigma_3 A_j + \sigma_4 H_i{*}S_j + \sigma_5 H_i{*}A_j + \sigma_6 X_i + \sigma_7 Y + \sigma_8 G + \varepsilon_2$$

where $\hat{O}_i$ is the predicted incremental length of stay due to adverse events experienced by a patient estimated in equation *(1)*, H_i is an indicator of whether a patient has high-risk condition *i;* S_j and A_j are indicators for specialist and avoider physicians, respectively (with generalist physicians as the reference category); X_i is a vector of characteristics of the patient and her delivering physician; and G is a vector of group practice fixed effects. The vector X_i includes an indicator of whether the physician is female, years of physician experience and its quadratic, an indicator if experience is missing, an indicator for New York patients, calendar year indicators, a patient's age and its quadratic, an indicator if age is missing, an indicator is a patient is a minority (black or Hispanic) and an indicator if race/ethnicity is missing, indicators for whether a patient is uninsured or covered by Medicaid, an indicator for whether the patient has had a previous cesarean section, and indicators for each of the other 11 high-risk conditions. It is worth repeating that we classify physicians as specialists, generalists or avoiders on their weekday deliveries,

[17] As a robustness check, we also create a dichotomous outcome variable measuring whether a woman experienced any of the major adverse outcomes, where major is defined as an outcome with an incremental effect on length of stay (ϕ_1) of 0.5 or greater. The overall pattern of results with this dependent variable was similar.

when intentional matching occurs, whereas we examine the impact of matching due to within-group differentiation on outcomes based on weekend deliveries, where our simulation has validated the assumption that the call schedule randomly assigns high- and low-risk patients to physicians within the practice.

We rely on linear models to test the effects of matching on mothers' outcomes. Because the dependent variable is generated from the predicted coefficients from equation *(1)*, we correct for heteroskedasticity by weighting each observation in equation *(2)* by the inverse of the standard error of the prediction of $\hat{O}_i$ (Mayda 2006).[18] Robust standard errors are calculated, however no clustering is done: despite the hierarchical nature of the data, the intra-group correlation coefficients of the regression residuals are all zero, indicating that clustering is unnecessary and inefficient, particularly with the large number of patients per group.[19]

If the following hold: (1) assigning a patient who has a particular (observed) health condition to a physician within the practice who specializes in that condition (i.e., matching) improves the patient's delivery outcome, (2) assigning such a patient to an avoider (i.e., mismatching) worsens her outcome, (3) assigning a patient who does not have a particular health condition to a physician within the practice who avoids that condition (i.e., matching) improves the patient's delivery outcome, and (4) assigning such a patient to a specialist (i.e., mismatching) worsens her outcome, then we would expect the following results:

- $\sigma_2 > 0$: low-risk patient outcomes are worse when randomly assigned to a specialist physician (i.e., a mismatch) rather than a generalist physician (the omitted category)
- $\sigma_3 < 0$: low-risk patient outcomes are better when randomly assigned to an avoider (i.e., a match) rather than a generalist
- $\sigma_3 - \sigma_2 < 0$: low-risk patient outcomes are better when randomly assigned to an avoider rather than a specialist
- $\sigma_4 < 0$: high-risk patient outcomes are better when randomly assigned to a specialist (i.e., a match) rather than a generalist
- $\sigma_5 > 0$: high-risk patient outcomes are worse when randomly assigned to an avoider (i.e., a mismatch) rather than a generalist

[18] The pattern of results from these weighted least squares estimates was not different from the pattern of estimates using robust standard errors alone to account for heteroskedasticity.
[19] Hansen (2007) provides additional details.

- $\sigma_4 - \sigma_5 < 0$: high-risk patient outcomes are better when randomly assigned to a specialist rather than an avoider.

In Table 4 we report estimated coefficients and robust standard errors separately for a woman's delivery outcome if she has one of 10 specific health conditions, as well as whether she has any high-risk condition. Overall, the results provide some evidence that matching improves mothers' outcomes and mismatching harms them. Of the 66 tests of the hypotheses listed above (i.e., six tests in 11 regressions), the hypotheses are accepted at the five-percent level in seven instances (10.6 percent), while providing statistically significant counterfactual evidence in only one (1.5 percent) instance, indicating that the counterfactual findings are likely due to chance while the supporting results are not

The benefits of matching occur primarily from assigning low-risk patients to avoider physicians as opposed to generalist or specialist physicians. Specifically, matching low-risk patients with avoider physicians ($\sigma_3 < 0$, or $\sigma_3 - \sigma_2 < 0$) significantly (at the five-percent level) improves outcomes for women who do not have any high-risk condition, any hypertension, a malpositioned fetus, or antepartum bleeding. Surprisingly, there is only one instance (for an unengaged fetal head) of a statistically significant (at the five-percent level) benefit from matching high-risk patients with specialists rather than generalists ($\sigma_4 < 0$), and even here the difference between a specialist and avoider is insignificant. Contrary to the hypotheses above, in one case there is a statistically significant benefit associated with mismatching: assigning a woman with any hypertension to a physician who avoids that condition rather than a specialist is associated with improved outcomes, although being assigned to an avoider is not statistically different from a generalist. Finally, there is only one case where mismatching has negative health consequences for the mother: assigning a woman who does not have antepartum bleeding to a specialist rather than a generalist ($\alpha_2 > 0$) is associated with worse outcomes.[20]

[20] Mismatching a pregnancy without unengaged fetal head to a specialist also has worse outcomes at the ten-percent level.

When they exist, the positive benefits of matching and the negative consequences of mismatching are large. Based on the mean of the dependent variable for women who do not have the specific health condition, there are large predicted improvements in a mother's delivery outcome if she is matched with an avoider physician rather than a specialist for any hypertension (22 percent fewer predicted days in the hospital due to complications), malpositioned fetus (23 percent), or antepartum bleeding (52 percent). Similarly, mothers without any of the 12 risk factors experience 19 percent fewer predicted hospital days due to complications when paired with an avoider physician who treats few patients with any high-risk factors rather than a generalist. In the one instance of negative consequences of mismatching, a woman who does not have antepartum bleeding is predicted to spend 41 percent more days in the hospital due to complications if assigned to a physician who specializes in that risk factor relative to being assigned to a generalist.

Our approach measures the effect of within-group specialization on the outcomes of women who go into labor and deliver on the weekend. The results are likely to underestimate the overall benefits of specialization and matching and underestimate the costs associated with mismatching for several reasons. First, groups might respond to the potential risks of mismatches for certain patients and assign them to specific physicians by inducing or scheduling their deliveries during the week, as supported by the observed differences in the health conditions of weekday and weekend patients (Table 1). If groups do this for the patients who benefit the most from matching, then the weekend estimates, which rely on random assignment for the remaining patients, should be lower-bound estimates of the benefits of non-random matching on health characteristics. That is, we estimate the benefits of specialization and matching on a relatively healthy group of patients who may benefit less from matching than the average patient. This also holds true if matching occurs across groups for conditions that are known at the outset of the woman's pregnancy. Second, we consider matching only along a single dimension (either a single risk factor, or an indicator for the presence of any risk factor) at a time, but the health outcomes might depend on successful matching across multiple dimensions simultaneously.[21] Third, we observe matching only for the delivery itself, whereas matching for prenatal care might also be important.[22] Fourth, because of data limitations, we do not observe the delivery and subsequent health outcome of the baby, who may be the primary beneficiary of specialization and matching.

The benefits of matching in Table 4 occur primarily from matching low-risk patients (i.e., patients who do not have a specific health condition) with avoider physicians. This raises the question of whether group practices are better able to facilitate such matches relative to solo practitioners for weekday deliveries, when the greatest opportunity for within-group referrals exists. To address this question we replicated Table 3 for low-risk instead of high-risk patients. As shown in Table 5, low-risk patients in differentiated group practices are generally more likely to be matched with an avoider physician than in solo or homogeneous group practices. Differentiated groups have higher matching for 10 of the 12 risk factors individually, although homogenous practices have a higher matching rate for the summary measure. Differentiated groups are significantly more likely to match low-risk patients with avoider physicians for antepartum bleeding (20% vs. 9%, $p<0.001$) and a malpositioned fetus (23% vs. 19%, $p<0.001$), the conditions with the largest benefits of matching. For the other two risk factors where matching significantly affected mothers' health (any hypertension and unengaged fetal head), however, homogeneous practices achieved greater matching. Similarly, low-risk patients in differentiated groups are much more likely to be mismatched with a specialist for most conditions

[21] However, patient risk factors do not correlate highly. Only two pairs had correlations of at least 0.10 (multiples and malpositioned fetus, $\rho=0.10$, and multiples and preterm gestation, $\rho=0.13$).
[22] In order to address this issue, we created an indicator variable if there is a specialist (and separately for an avoider) for a particular health condition in the patient's practice, regardless of whether that specialist actually delivered the baby on the weekend. These variables were interacted with the vector of dummies for the patient's health condition. Results from these models for the effects of matching for the delivery were qualitatively similar to those reported in Table 4, and they indicated little additional effect due to the potential to match for prenatal care.

including for antepartum bleeding (14% vs. 8%, p<0.001), where these mismatches resulted in worse outcomes for mothers.[23]

V. Matching on Style and Skill

A. Within-group matching by physician cesarean section style and unobserved patient preferences

There might be benefits of specialization beyond those based on a patient's observed health condition and a physician's experience treating that condition. In this section we examine whether group practices facilitate two other types of matching: directing patients who prefer to be treated aggressively to physicians who have such treatment styles; and directing patients with those preferences or high-risk patients to physicians who are skilled in a particular delivery method (cesarean sections).

Previous research suggests that obstetricians differ substantially in their treatment style, defined as their propensity to perform a cesarean section conditional on a patient's pre-delivery health condition, although these measures might be biased due to patient-physician matching on unobserved characteristics (Grant 2005; Epstein and Nicholson 2005). If patients have homogeneous preferences for this dimension of 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. Although we are not aware of any research that has examined preferences for cesarean sections specifically, mothers have strong, heterogeneous preferences for related factors in obstetrics, such as the site of delivery (e.g., Donaldson, Hundley, and Mapp 1998; Miller 2006). This suggests that some of the observed differentiation in physicians' treatment styles might arise when women with different preferences match with physicians who have particular styles. Such differentiation would improve welfare (Wolinsky 1993).

[23] Differentiated groups were actually less likely to mismatch pregnancies with unengaged fetal head (15% vs. 17%, p<0.001), which also had marginally significant negative health effects from such mismatching.

We use the same data set to examine whether group practices with differentiated physicians promote the matching of patients based on obstetricians' style and skill. Our measure of an obstetrician's style is his risk-adjusted cesarean section rate for patients randomly paired with him: that is, the probability that a patient who labors on the weekend would receive a cesarean section if treated by the physician, controlling for the mother's observed characteristics. A cesarean section is a surgical procedure and, all else equal, is more expensive than a vaginal delivery, requires a longer recovery period, and presents increased health risks for the mother. In certain clinical situations, however, a cesarean section improves the expected health of the mother and infant.

The cesarean section rates in Florida and New York have increased sharply since the mid-1990s. As of 2004, cesarean sections accounted for 34.4 percent of all deliveries in Florida, as displayed in Figure 1. Cesarean sections either can be scheduled in advance of labor or can occur after a woman goes into labor. The recent increase in the Florida cesarean section rate has been driven primarily by a doubling of the scheduled cesarean section rate, from 10.9 percent in 1996 to 24.8 percent of all deliveries in 2004. Scheduled cesarean sections now account for about two-thirds of all cesarean sections in Florida. Figure 1 also shows the percentage of total deliveries that are scheduled cesarean sections separately for women with and without a high-risk condition.[24] The latter, sometimes referred to as "patient-choice" cesarean sections, now constitute slightly more than one-half of all scheduled cesarean sections in Florida. Figure 2 shows comparable data on deliveries at in New York hospitals. Although cesarean sections are used less frequently in New York than in Florida, the basic patterns in cesarean section rates over time are similar between the two states.

To generate each physician's treatment style (cesarean section propensity), we perform cross-sectional ordinary least squares regressions for all patients, separately for Florida and New

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

York and separately for weekend deliveries for mothers who went into labor and weekday deliveries (Baicker, Buckles, and Chandra 2006):

$$(3)\quad C_{is} = \beta_j J + \beta_{ps} X_i + \beta_{ys} Y + \varepsilon_{3s}$$

$$(4)\quad C^d_{is} = \beta^d_j J + \beta^d_{ps} X_i + \beta^d_{ys} Y + \varepsilon_{4s}$$

C_{is} is an indicator variable for whether a patient i in state s received a cesarean section; J is a full vector of physician fixed effects; X is a vector of patient characteristics that affect the likelihood of receiving a cesarean section (age, race, indicators for whether age or race were missing, type of health insurance, and health conditions); Y is a vector of fixed effects for year of delivery; and the superscript d refers to weekday deliveries. The coefficient estimates of the patient characteristics from equations *(3)* and *(4)* are reported in Appendix B. From these estimates we generate a predicted propensity to receive a cesarean section for each patient, conditional on the patient's characteristics and holding fixed the physician's treatment style: $\hat{C}_{is} = \hat{\beta}_j J^* + \hat{\beta}_{ps} X_i + \hat{\beta}_{ys} Y_i$ for weekend laboring patients in state s, and $\hat{C}^d_{is} = \hat{\beta}^d_j J^* + \hat{\beta}^d_{ps} X_i + \hat{\beta}^d_{ys} Y_i$ for weekday patients, where J^* represents a physician selected arbitrarily from J. The overall median predicted cesarean section propensity was 0.140.

Among group practices that assign their weekend patients randomly, Equation 3 allows us to measure group members' treatment styles as if they were to treat an identical patient population. Specifically, the coefficients on the physician indicators, $\hat{\beta}_j$, in the weekend regressions measure the average adjusted probability that physician j performed a cesarean section on a patient randomly-assigned within the group practice, controlling for observed patient characteristics. Thus differences in $\hat{\beta}_j$ within a given medical group should be due to differences in style, i.e., physicians' perceptions about how to treat a common set of patients who labor, rather than due to within-practice physician-patient matching based on unobservable characteristics. Therefore, only the 407 physicians in group practices that assign their weekend patients randomly based on patients' predicted cesarean section propensity are included in the following analyses, where we use the simulation method described in Section III to identify practices that randomize.[25]

Because the randomness of the call schedule can eliminate unobserved differences in patient characteristics only across physicians in a given group, we repeat the estimates of equations *(3)* and *(4)* but replace the physician fixed effects with group practice fixed effects. With these results, we then generate the group-demeaned measures of each physician's c-section style, $\left(\hat{\beta}_j - \hat{\beta}_g\right)$, where the subscript *g* refers to a group practice. We find a degree of within-group variation across physicians' c-section style, spanning -0.024 (i.e., 2.4 percentage points) at the 10th percentile to 0.021 at the 90th percentile, with a mean of -0.001 and a standard deviation of 0.019.

With measures of physicians' treatment styles, we consider whether patients match with physicians whose styles are consistent with their preferences. For example, a patient who places a high cost on the adverse outcomes associated with a cesarean section could reduce the probability she will have a cesarean section by scheduling her labor to be induced by a conservative obstetrician with a low cesarean section style, or she may request that a particular physician perform the delivery if she goes into labor during the week. Conversely, a woman in good health who wants to avoid laboring altogether might be matched to a physician in the practice willing to perform a patient-choice cesarean section.[26]

If patients with particular preferences match within groups with physicians who have commensurate treatment styles, patients who want intensive treatment would match with physicians who have a large/positive $\hat{\beta}_j$ relative to their colleagues within the group practice

[25] One limitation of investigating within-firm heterogeneity and matching in this context is that many of the firms are small: 45% of these physicians are in groups of 2, and the mean number of physicians in the group is 3.2.

[26] 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" (Brody 2003).

$\left(\hat{\beta}_j - \hat{\beta}_g\right)$, and patients who want less intensive treatment would match with physicians who have a small/negative $\hat{\beta}_j$. As a result, a physician who performs many patient-choice scheduled cesarean sections on weekdays, for example, would have a large $\left(\hat{\beta}_j^d - \hat{\beta}_g^d\right)$ because his actual proportion of cesarean 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.

As a result, if group practices promote matching based on patient preferences and physician styles, there would be more variation across physicians in a group in their weekday cesarean section rates when there are ample opportunities for patients and physicians to intentionally match than in their weekend rates: $\sigma^d_g/\sigma_g > 1$ where σ^d_g = variance($\beta^d_j \mid g$).[27] This should be true even when cesarean section rates are adjusted for observed patient characteristics, because patients with unobserved preferences will be matching with specific physicians. Physician treatment styles on the weekend are measured less precisely than weekday styles because the weekend sample size is substantially smaller. Therefore, we report the results of the variance test both without and with a Bayesian shrinkage adjustment. The Bayesian shrinkage adjustment will move a physician's treatment style toward the sample mean if it has a low signal-to-noise ratio (Hofer et al. 1999), which may result in less variation between physicians on the weekend (due to a relatively small sample size) relative to the weekday.

Kolmogorov-Smirnov tests reject normality of the distributions of these group-level ratios of the weekday and weekend variances in physician cesarean section style. Consequently, we report results from the non-parametric Wilcoxon signed-rank tests in Panel A of Table 6. These reject the hypothesis that $\sigma^d_g > \sigma_g$. We confirm this by testing whether the median group's ratio of the weekend to weekday intra-group variances is equal to one, where we bootstrap the

[27] We would also expect σ^d_g to exceed σ_g if patients with unobserved health characteristics that increase the expected marginal benefit of cesarean sections to match with physicians who have a high propensity to perform cesarean sections to improve their expected health outcome.

standard errors with 1,000 iterations. Results are similar when we consider only groups of three or more physicians. These results are not consistent with sorting based on unobserved patient preferences for physician treatment styles.

One shortcoming of testing whether the weekday and weekend variances are equal is that the variables are affected by both physician and patient decisions. For example, physicians with high cesarean section styles might convince some of their patients that do not have strong preferences for cesarean sections to schedule cesarean sections during the week. If this phenomenon accounts for the similar variance between weekday and weekend treatment styles, the welfare benefits are more ambiguous than they would be if the differences are driven by well-informed patients.[28]

We therefore also test whether the correlation $\left(\hat{\beta}_j - \hat{\beta}_g, Q_j^d - Q_g^d\right)$ is greater than zero, where $(Q^d_j - Q^d_g)$ is the difference between the number of weekday scheduled patient-choice cesarean sections physician *j* performs and the average of his group practice. Finding that physicians who perform more cesarean sections relative to their colleagues on randomly-assigned laboring weekend patients also perform a large number of scheduled weekday patient-choice cesarean sections would suggest that patient preferences are driving the matches. Scheduled high-risk cesarean sections, on the other hand, could be initiated by either a patient or a physician.

In Panel B of Table 6 we report the correlations of $\left(\hat{\beta}_j - \hat{\beta}_g, Q_j^d - Q_g^d\right)$ for all cesarean sections and each type of cesarean section separately. The correlations are significantly different from zero for all cesarean sections and for both high-risk and patient-choice cesarean sections. The latter are presumably initiated by the patient. Low-risk patients who nevertheless want an

[28] The ambiguity results from the fact that physicians with higher cesarean section propensities might also have greater skill for cesarean sections than vaginal deliveries, thus leading to better outcomes (Chandra and Staiger 2007), but this could be offset by the mothers' preferences for a vaginal delivery. However, the correlation of physicians' cesarean section style and skill measures, relative to their peers', is 0.05 (p=0.32), where the positive coefficient indicates that those with worse cesarean section outcomes tend to perform more cesarean sections on their randomly-paired weekend laboring patients. This low correlation allows us to separately test matching based on style and skill, which we investigate in the analysis that follows.

elective cesarean section appear to match with a physician in the practice who has a style inclined toward cesarean sections. This supports the interpretation that patient preferences create some of the observed variation in obstetricians' treatment styles. Physicians in the sample performed an average of 13.6 weekday patient-choice cesarean sections per year. A physician who is at the 90th percentile of the difference between his own weekend cesarean section rate and that of his group is predicted to perform 8.8 more weekday patient choice cesarean sections per year than his colleague who is at the 10th percentile of the distribution.[29] This represents a fairly large (65 percent) increase in a physician's annual quantity of patient-choice cesarean sections. The apparent discrepancy between Panels A and B of Table 6 in whether unobserved patient preferences contribute to weekday matching is due to the fact that physicians with styles geared toward cesarean sections also perform a higher total volume of weekday deliveries relative to their peers (ρ=0.12, p=0.019). The result is that these physicians' weekday cesarean section volumes are higher but their weekday styles are not farther from their group's average , because the definition of style depends on the total volume as well.

B. Within-group matching by physician cesarean section skill

We also test whether group practices match physicians and patients based on skill and patients' clinical characteristics and/or their unobserved preferences. Specifically, we first examine whether a groups' patients are matched with physicians during the week based on a physician's absolute or comparative advantage in performing cesarean section deliveries, relative to his peers in his group, and the appropriateness of a c-section for the patient, predicted from equation *(4)* above. To examine this, we estimate:

$$(5)\quad \hat{O}_i = \xi_{0m} + \xi_{1m} X_i + \xi_{jm} J + \varepsilon_{5m} \ ,$$

[29] To calculate this, we regressed $(Q^d_j - Q^d_g)$ on $(\beta_j - \beta_g)$ and used the estimated coefficients to generate predicted values of $(Q^d_j - Q^d_g)$ at the 10th and 90th percentiles of $(\beta_j - \beta_g)$. At the 10th percentile of $(\beta_j - \beta_g)$ (-0.027), the predicted value of $(Q^d_j - Q^d_g)$ was -4.45, and at the 90th percentile of $(\beta_j - \beta_g)$ (0.024), the predicted value of $(Q^d_j - Q^d_g)$ was 4.31.

separately by mode of delivery *m* for laboring weekend patients of groups that randomly assign their weekend patients to physicians. $\hat{O}_i$ is a woman's delivery outcome generated from equation *(1)* above; *X* is a vector of patient characteristics; *J* is a full set of physician indicator variables; ξ_{jc} is the physician's skill at performing cesarean sections; and ξ_{jv} is his skill at performing vaginal deliveries. Lower values of $\hat{O}_i$ indicate better outcomes. A physician will have a negative ξ_{jc} if his patients who deliver by cesarean section have fewer or less severe complications than expected given the patients' observed pre-delivery health. Because we rely on the random within-group pairing of physicians with patients who labor on the weekend, as before we restrict the analysis to those groups that randomly assign patients based on clinical factors. Because this identification strategy eliminates unobserved differences across physicians in the same group only, we continue to examine group-demeaned measures of skill by repeating equation *(5)* to generate $\left(\xi_{jc} - \xi_{gc}\right)$.

Physicians vary substantially in their skill at performing cesarean sections. For example, weekend patients randomly assigned to physicians at the 10th percentile of the distribution of the absolute advantage measure of skill average 0.20 more hospital days due to complications than the group's average than those treated by physicians at the 90th percentile, and the standard deviation of this distribution is 0.10.

To consider matching based on absolute advantage and the appropriateness of a cesarean section for a patient, we test whether the correlation $\left(\xi_{jc} - \xi_{gc}, \hat{C}^d_{js} - \hat{C}^d_{gs}\right)$ is less than zero. Results in Table 7 Panel A indicate that physicians with better weekend outcomes treat patients for whom cesarean deliveries are more appropriate during the week, supporting the hypothesis that weekday patients match with physicians based on skill and observed clinical characteristics. This is consistent with group practices internally referring the most difficult cesarean section cases to physicians who are skilled at performing cesarean sections. We also consider matching based on absolute advantage and patient preferences by examining whether the correlation $\left(\xi_{jc} - \xi_{gc}, \beta_{js}^{d} - \beta_{gs}^{d}\right)$ is less than zero. The term $\left(\beta_{js}^{d} - \beta_{gs}^{d}\right)$ is the physician's group-demeaned, risk-adjusted weekday cesarean section rate, which reflects differences in patients' preferences. Results in Table 7 suggest that this type of matching occurs.

Alternative interpretations exist for both of these measures of matching based on skill. Both correlations would also be negative if skill is acquired by learning through performing more cases. Likewise the second correlation could be negative if physicians more highly skilled at cesarean sections are more likely to perform them, holding observed and unobserved patient characteristics fixed. Neither of these alternatives is supported by the correlation ($\xi_{jc} - \xi_{gc}$, $Q^{d}_{jc} - Q^{d}_{gc}$), which indicates that physicians who perform more weekday patient-choice cesarean sections did not have better weekend outcomes (ρ=-0.056, P=0.256). Likewise, the negative and insignificant relationship between physicians' cesarean section skill and style reported above does not support the alternative explanation that physician treatment decisions rather than unobserved patient preferences account for the higher propensity for physicians more highly skilled at cesarean sections to also have higher weekday risk-adjusted cesarean section rates.

Because absolute advantage in cesarean section skill ignores opportunity costs, we repeat this analysis using physicians' comparative advantage for cesarean sections over vaginal deliveries relative to his peers, defined as $\left(\frac{\xi_{jc}}{\xi_{jv}} - \frac{\xi_{gc}}{\xi_{gv}}\right)$. Results in Panel B of Table 7 are similar to those in Panel A but smaller in size and significance. This greater importance of absolute advantage is consistent with existing theory, which predicts that absolute advantage will be a better proxy for optimal assignment than comparative advantage (MacDonald and Markusen 1985, Sattinger 1993.) Together these results suggest that groups match patients to physicians based on their skill in welfare-increasing ways.

VI. Conclusion

In this paper we observed a professional services market in which firms play an important role in coordinating differentiated high-skilled workers due to barriers in markets that derive from imperfect or asymmetric information, incentives, and regulation. We found that the greatest amount of additional matching due to firm coordination occurred in situations where regulations and payment methods created the highest financial barriers to across-firm referrals. Specifically, obstetrics partnerships appear to facilitate greater matching by promoting within-group referrals and pairing specialized physicians with high-risk patients, particularly for health conditions that become known only later in the pregnancy. In addition to matching based on patient and physician clinical characteristics, we found some evidence of matching within groups along other dimensions. Internally-differentiated partnerships appropriately matched patients with observed clinical characteristics and unobserved preferences to physicians based on physicians' absolute skill advantage. However, the extent to which partnerships promoted matching based on unobserved patient preferences and physician style was ambiguous.[30] Despite the institutional peculiarities of obstetrics markets, these results are consistent with a number of related theoretical and empirical studies in other settings. Additional work could consider whether matching in this context is also influenced by organizational characteristics other than the degree of specialization, or the presence of a managerial hierarchy (Garicano and Hubbard, 2007).

Partnerships of heterogeneous obstetricians and subsequent patient-physician matching enhances welfare via both improved maternal health outcomes and greater incorporation of patient preferences. Yet it remains unclear whether the overall level of differentiation across physicians in obstetrics markets is optimal. Such issues are important to consider as various technologies are developed to standardize physician treatment decisions. Similarly, from the perspective of the industrial organization of and labor force planning for obstetrics, it remains

[30] The identification strategy precluded a direct comparison between the degree of matching in groups with differentiated physicians with the degree in homogeneous groups and solo practices.

unclear whether greater matching and welfare would result from more within-firm heterogeneity, more consolidation of physicians into larger partnerships, or expanded scope of existing partnerships (Farrell and Scotchmer 1988).

One reason for this ambiguity is due to the random pairing of patients and physicians within obstetrics partnerships. While econometrically convenient to generate unbiased measures of physicians' styles and skills, it also incorporates a tradeoff into partnerships' decisions about the degree of internal differentiation. Referring patients between physicians allows a group practice to match patients with appropriate physicians during the week, but this also creates mismatches from patients and physicians being paired randomly due to call schedules on week nights and weekends. We found that differentiated groups were less likely than solo practices to have "avoider" physicians who treat fewer high-risk patients than expected, and more likely to match low-risk patients with specialists. This random pairing of patients and physicians for deliveries also might results in less heterogeneity than in other high-skilled labor markets. Alternatively, the non-emergent nature of pregnancy, as opposed to the delivery itself, likely contributes to more matching relative to other high-skilled labor markets, particularly in medicine.

References

Ackerberg, Daniel A., and Maristella Botticini. 2002. "Endogenous Matching and the Empirical Determinants of Contract Form." *Journal of Political Economy,* 110(3): 564-591.

Afendulis, Christopher C., and Daniel P.Kessler. 2007. "Tradeoffs from Integrating Diagnosis and Treatment in Markets for Health Care." *The American Economic Review*, 97(3): 1013-1020.

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *The American Economic Review*, 62(5): 777-795.

Baicker, Katherine, Kasey S.Buckles, and Amitabh Chandra. 2006. "Geographic Variation in the Appropriate Use of Cesarean Delivery." *Health Affairs*, 25: w355-w367.

Bolton, Patrick, and Mathias Dewatripont. 1994. "The Firm as a Communication Network." *The Quarterly Journal of Economics*, 109(4): 809-839.

Britt, David W., Robert D. Eden, and Mark I. Evans. 2006. "Matching risk and resources in high-risk pregnancies." *The Journal of Maternal-Fetal and Neonatal Medicine*, 19(10): 645-650.

Brody, Jane E. 2003. "As Cases of Induced Labor Rise, So Do Experts' Concerns." *The New York Times*, Section F Col. 1, p. 7.

Chandra, Amitabh and Douglas Staiger. 2007. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy*, 115(1): 103-140.

Chou, Eric S. 2007. "The Boundaries of Firms as Information Barriers." *The Rand Journal of Economics*, 38(3): 733-746.

Donaldson, Cam, Vanora Hundley, and Tracy Mapp. 1998. "Willingness to Pay: A Method for Measuring Preferences for Maternity Care?" *Birth*, 25(1): 32-39.

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

Gaynor, Martin, and Paul Gertler. 1995. "Moral Hazard and Risk Spreading in Partnerships." *The Rand Journal of Economics*, 26(4): 591-613.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *The Journal of Political Economy*, 108(5): 874-904.

Garicano, Luis and Thomas N. Hubbard. 2007. "Managerial Leverage is Limited by the Extent of the Market: Hierarchies, Specialization, and the Utilization of Lawyers' Human Capital." *Journal of Law and Economics*, 50: 1-43.

Garicano, Luis and Tano Santos. 2004. "Referrals." *The American Economic Review*, 94(3): 500-525.

Grant, Darren. 2005. "Information and Sorting in the Market for Obstetrical Services." *Health Economics*, 14: 703-719.

Gregory, Kimberly D., Lisa M. Korst, Jeffrey A. Gornbein, and Lawrence D. Platt. 2002. "Using Administrative Data to Identify Indications for Elective Primary Cesarean Delivery." *Health Services Research,* 37(5): 1387-1401.

Hansen, Christian B. 2007. "Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects." *Journal of Econometrics*, 2: 670-694.

Hart, Oliver and John Moore. 2005. "On the Design of Hierarchies: Coordination versus Specialization." *Journal of Political Economy*, 113(4): 675-702.

HealthGrades. 2007. *Hospital Report Cards^TM Maternity Care and Women's Health 2007-2008 Methodology White Paper (2003- 2005 Data)*, http://www.healthgrades.com/consumer/index.cfm?fuseaction=modnw&modtype=content&modact=hrc_methodology_consumer, accessed April 25, 2008.

Henry, Olivia A., Kimberly D. Gregory, Calvin J. Hobel, and Lawrence D. Platt. 1995. "Using ICD-9 Codes to Identify Indications for Primary and Repeat Cesarean Sections: Agreement With Clinical Records." *American Journal of Public Health,* 85(8): 1143-1145.

Hoerger, Thomas J., and Leslie Z.Howard. 1995. "Search Behavior and Choice of Physician in the Market for Prenatal Care." *Medical Care*, 33(4): 332-349.

Hofer, Timothy P., Rodney A. Hayward, Sheldon Greenfield, Edward H. Wagner, Sherrie H. Kaplan, and Willard G. Manning. 1999. "The Unreliability of Individual Physician 'Report Cards' for Assessing the Costs and Quality of Care of a Chronic Disease." *JAMA*, 281(22): 2098-2105.

Hubbard, Thomas N. 1998. "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market." *The RAND Journal of Economics*, 29(2): 406-426.

Levin, Jonathan, and Steven Tadelis. 2005. "Profit Sharing and the Role of Professional Partnerships." *The Quarterly Journal of Economics,* 120(1): 131-171.

MacDonald, Glenn M. 1982. "Information in Production." *Econometrica*, 50(5): 1143-1162.

MacDonald, Glenn M., and James R. Markusen. 1985. "A Rehabilitation of Absolute Advantage." *The Journal of Political Economy*, 93(2): 277-297.

Martin, Joyce A., Brady E. Hamilton, Paul D. Sutton, et al. 2003. "Births: Final Data for 2002." *National Vital Statistics Report*, 52(10): 1-116.

Mayda, Anna M. 2006. "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants." *Review of Economics and Statistics*, 88(3): 510-530.

Miller, Amalia R. 2006. "The Impact of Midwifery-Promoting Public Policies on Medical Interventions and Health Outcomes." *Advances in Economic Analysis and Policy,* 6(1): 1-34.

Morrison, Alan D, and William J. Wilhelm, Jr. 2004. "Partnership Firms, Reputation, and Human Capital." *The American Economic Review*, 94(5): 1682-1692.

Sattinger, Michael. 1993. “Assignment Models of the Distribution of Earnings.” *Journal of Economic Literature*, 31(2): 831-880.

Schauberger, Charles W., Robert K. Gribble, and Brenda L.Rooney. 2007. “On Call: A Survey of Wisconsin Obstetric Groups.” *The American Journal of Obstetrics and Gynecology*, 196(39): 39.e1-39.e4.

Wolinsky, Asher. 1993. “Competition in a Market for Informed Experts’ Services.” *The RAND Journal of Economics*, 24(3): 380-398.

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

Variable	Weekend Laboring Deliveries Mean	Weekend Laboring Deliveries Standard Deviation	All Weekday Deliveries Mean	All Weekday Deliveries Standard Deviation
Patient Health Condition				
Any of the 12 high-risk conditions	0.173	0.378	0.262	0.440
Preterm gestation	0.065	0.246	0.070	0.256
Any hypertension	0.041	0.198	0.068	0.252
Malpositioned fetus	0.028	0.165	0.068	0.251
Macrosomia	0.015	0.120	0.029	0.167
Maternal soft tissue disorder	0.015	0.120	0.028	0.165
Pregnancy-related severe hypertension	0.004	0.066	0.012	0.107
Antepartum bleeding	0.009	0.092	0.017	0.128
Multiples (twins)	0.006	0.075	0.013	0.112
Herpes	0.010	0.097	0.014	0.117
Unengaged fetal head	0.011	0.105	0.014	0.119
Uterine scarring	0.000	0.020	0.002	0.046
Congenital anomaly or chromosomal abnormality	0.001	0.024	0.001	0.032
N Deliveries	312,695		1,327,461	
Patient Demographics				
Age	27.1	6.2	27.8	6.3
Black or Hispanic	0.512	0.500	0.475	0.499
Medicaid or other public health insurance	0.513	0.500	0.471	0.499
Florida	0.480	0.500	0.534	0.499
Delivery Characteristics				
Mother went into labor	1.000	0.000	0.789	0.408
Delivery performed by c-section:	0.099	0.299	0.302	0.459
- Scheduled, high-risk	0.000	0.000	0.104	0.305
- Scheduled, patient-choice	0.000	0.000	0.107	0.309
- Unscheduled (following labor)	0.099	0.299	0.091	0.288
Health outcome (continuous measure)*	0.030	0.169	0.039	0.216
Predicted C-section probability**	0.077	0.120	0.335	0.295

* Generated from regressions reported in Appendix A
** Generated from regressions reported in Appendix B

Table 2: Extent of Specialization on Weekdays, by Health Condition and Practice Type

		Practice Type						Mean
Patient Health Condition	Overall (N=1,655)	Solo (N=720)	Group (N=945)	Group - Solo	P-value	Homo-geneous Group	Differen-tiated Group	Deliveries with Condition per MD, 1999-2004
Any of the 12 high-risk conditions						N=347	N=598	
Specialists	0.20	0.19	0.20	0.01	0.621	0.16	0.23	371.2
Avoiders	0.28	0.33	0.24	-0.09	0.000	0.23	0.25	207.7
Preterm gestation						N=391	N=554	
Specialists	0.15	0.14	0.17	0.03	0.109	0.11	0.21	149.1
Avoiders	0.28	0.32	0.26	-0.06	0.010	0.20	0.30	44.8
Any hypertension						N=305	N=640	
Specialists	0.19	0.16	0.21	0.05	0.010	0.12	0.25	134.0
Avoiders	0.25	0.28	0.23	-0.06	0.009	0.19	0.25	47.5
Malpositioned fetus						N=439	N=506	
Specialists	0.13	0.13	0.13	0.00	0.771	0.05	0.20	120.9
Avoiders	0.16	0.19	0.14	-0.05	0.006	0.07	0.19	48.6
Macrosomia						N=366	N=579	
Specialists	0.17	0.18	0.16	-0.02	0.282	0.10	0.19	60.7
Avoiders	0.25	0.30	0.21	-0.09	0.000	0.22	0.20	13.3
Maternal soft tissue disorder						N=294	N=651	
Specialists	0.17	0.16	0.18	0.01	0.491	0.14	0.19	55.8
Avoiders	0.25	0.27	0.23	-0.04	0.087	0.18	0.26	12.8
Pregnancy-related severe hypertension						N=443	N=502	
Specialists	0.10	0.08	0.11	0.03	0.052	0.05	0.16	37.1
Avoiders	0.19	0.23	0.16	-0.07	0.000	0.07	0.23	4.0
Antepartum bleeding						N=483	N=462	
Specialists	0.09	0.09	0.08	-0.01	0.687	0.03	0.14	37.1
Avoiders	0.11	0.12	0.09	-0.03	0.096	0.00	0.19	7.1
Multiples (twins)						N=445	N=500	
Specialists	0.09	0.06	0.12	0.06	0.000	0.05	0.18	35.4
Avoiders	0.15	0.15	0.14	0.00	0.850	0.04	0.23	4.5
Herpes						N=371	N=574	
Specialists	0.13	0.12	0.14	0.03	0.098	0.06	0.19	35.1
Avoiders	0.18	0.21	0.17	-0.05	0.017	0.09	0.21	5.1
Unengaged fetal head						N=347	N=598	
Specialists	0.17	0.20	0.14	-0.05	0.004	0.11	0.17	34.0
Avoiders	0.25	0.22	0.27	0.05	0.022	0.17	0.33	4.0
Uterine scarring						N=239	N=706	
Specialists	0.07	0.08	0.07	0.00	0.732	0.04	0.08	7.5
Avoiders	0.32	0.32	0.33	0.01	0.521	0.23	0.36	0.1
Congenital anomaly or chromosomal abnormality						N=430	N=515	
Specialists	0.07	0.05	0.08	0.03	0.021	0.03	0.11	6.1
Avoiders	0.23	0.24	0.22	-0.03	0.223	0.14	0.29	0.1
Unweighted average								
Specialists	0.13	0.12	0.13	0.01		0.07	0.17	
Avoiders	0.22	0.24	0.20	-0.03		0.13	0.25	

Notes: Solo includes physicians who practiced alone and physicians in group practices with other physicians who failed to meet the minimum threshold of 100 weekend deliveries between 1999 and 2004.

Table 3: Extent of Matching and Mismatching for High-Risk Patients on Weekdays, by Health Condition and Practice Type

Patient Health Condition	Overall	Practice Type: Solo & Homogeneous Group	Practice Type: Differentiated Group	Practice Type: Differentiated Group - (Solo & Homogeneous Group)	P-value
Any of 12 risk factors					
Matched with Specialist	0.31	0.29	0.33	0.04	0.000
Mismatched with Avoider	0.24	0.26	0.21	-0.05	0.000
Preterm gestation					
Matched with Specialist	0.34	0.28	0.44	0.15	0.000
Mismatched with Avoider	0.18	0.19	0.18	-0.01	0.002
Any hypertension					
Matched with Specialist	0.35	0.30	0.44	0.14	0.000
Mismatched with Avoider	0.17	0.18	0.16	-0.02	0.000
Malpositioned fetus					
Matched with Specialist	0.26	0.18	0.41	0.23	0.000
Mismatched with Avoider	0.13	0.13	0.13	0.00	0.450
Macrosomia					
Matched with Specialist	0.39	0.36	0.44	0.08	0.000
Mismatched with Avoider	0.13	0.15	0.09	-0.05	0.000
Maternal soft tissue disorder					
Matched with Specialist	0.39	0.38	0.39	0.01	0.079
Mismatched with Avoider	0.12	0.12	0.14	0.02	0.000
Pregnancy-related severe hypertension					
Matched with Specialist	0.37	0.31	0.48	0.17	0.000
Mismatched with Avoider	0.07	0.08	0.06	-0.01	0.002
Antepartum bleeding					
Matched with Specialist	0.21	0.18	0.29	0.11	0.000
Mismatched with Avoider	0.05	0.03	0.08	0.04	0.000
Multiples (twins)					
Matched with Specialist	0.30	0.21	0.48	0.28	0.000
Mismatched with Avoider	0.05	0.04	0.08	0.04	0.000
Herpes					
Matched with Specialist	0.36	0.30	0.47	0.17	0.000
Mismatched with Avoider	0.07	0.07	0.07	0.00	0.342
Unengaged fetal head					
Matched with Specialist	0.43	0.43	0.41	-0.02	0.014
Mismatched with Avoider	0.07	0.05	0.11	0.06	0.000
Uterine scarring					
Matched with Specialist	0.31	0.30	0.33	0.03	0.065
Mismatched with Avoider	0.00	0.00	0.00	0.00	0.120
Congenital anomaly or chromosomal abnormality					
Matched with Specialist	0.47	0.43	0.54	0.11	0.000
Mismatched with Avoider	0.00	0.00	0.00	0.00	
Unweighted average					
Matched with Specialist	0.35	0.30	0.43	0.12	
Mismatched with Avoider	0.09	0.09	0.09	0.01	

Table 4: Effects of matching for delivery on the mothers' delivery outcomes, measured by incremental length of stay due to complications

	Any of 12 Risk Factors	Preterm gestation	Any hyper-tension	Mal-positioned fetus	Macro-somia	Maternal soft tissue disorder	Pregnancy-related severe hyper-tension	Ante-partum bleeding	Multiples	Herpes	Unen-gaged fetal head
Low risk patient * specialist (σ_2)	-0.003	-0.003	0.002	0.001	0.005	0.003	-0.009*	0.011***	0.002	-0.001	0.008*
	[0.002]	[0.002]	[0.003]	[0.003]	[0.003]	[0.004]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]
Low risk patient * avoider (σ_3)	-0.005***	-0.003	-0.004**	-0.006**	-0.000	-0.000	-0.001	-0.004	-0.001	-0.001	0.001
	[0.002]	[0.002]	[0.002]	[0.003]	[0.003]	[0.002]	[0.002]	[0.002]	[0.003]	[0.002]	[0.002]
High risk patient * specialist (σ_4)	0.004	0.015	0.016	0.016	0.007	0.024	0.107	0.013	-0.095	0.008	-0.094**
	[0.007]	[0.009]	[0.015]	[0.018]	[0.014]	[0.031]	[0.232]	[0.028]	[0.061]	[0.019]	[0.042]
High risk patient * avoider (σ_5)	-0.002	0.004	-0.012	0.020	0.043	-0.006	0.078	0.025	-0.098	0.042	-0.054
	[0.005]	[0.007]	[0.011]	[0.029]	[0.027]	[0.023]	[0.207]	[0.029]	[0.060]	[0.037]	[0.049]
N	41309	38032	39691	22899	28662	23174	21833	19600	16200	22318	23308
R^2	0.013	0.016	0.020	0.014	0.010	0.016	0.017	0.030	0.017	0.015	0.013
(σ_3) - (σ_2)	-0.003	0.001	-0.007	-0.007	-0.005	-0.003	0.008	-0.014	-0.003	0	-0.007
	[0.003]	[0.003]	[0.003]**	[0.004]*	[0.004]	[0.005]	[0.006]	[0.004]***	[0.005]	[0.004]	[0.005]
(σ_4) - (σ_5)	0.006	0.011	0.028	-0.004	-0.036	0.029	0.028	-0.012	0.003	-0.034	-0.041
	[0.007]	[0.010]	[0.013]**	[0.031]	[0.029]	[0.032]	[0.302]	[0.038]	[0.020]	[0.039]	[0.033]
Mean of dependent variable for high risk patients	0.0459	0.0289	0.0681	0.0567	0.0444	0.0716	0.190	0.0402	0.0650	0.0422	0.0760
Mean of dependent variable for low risk patients	0.0261	0.0289	0.0319	0.0309	0.0274	0.0267	0.0289	0.0268	0.0283	0.0300	0.0271

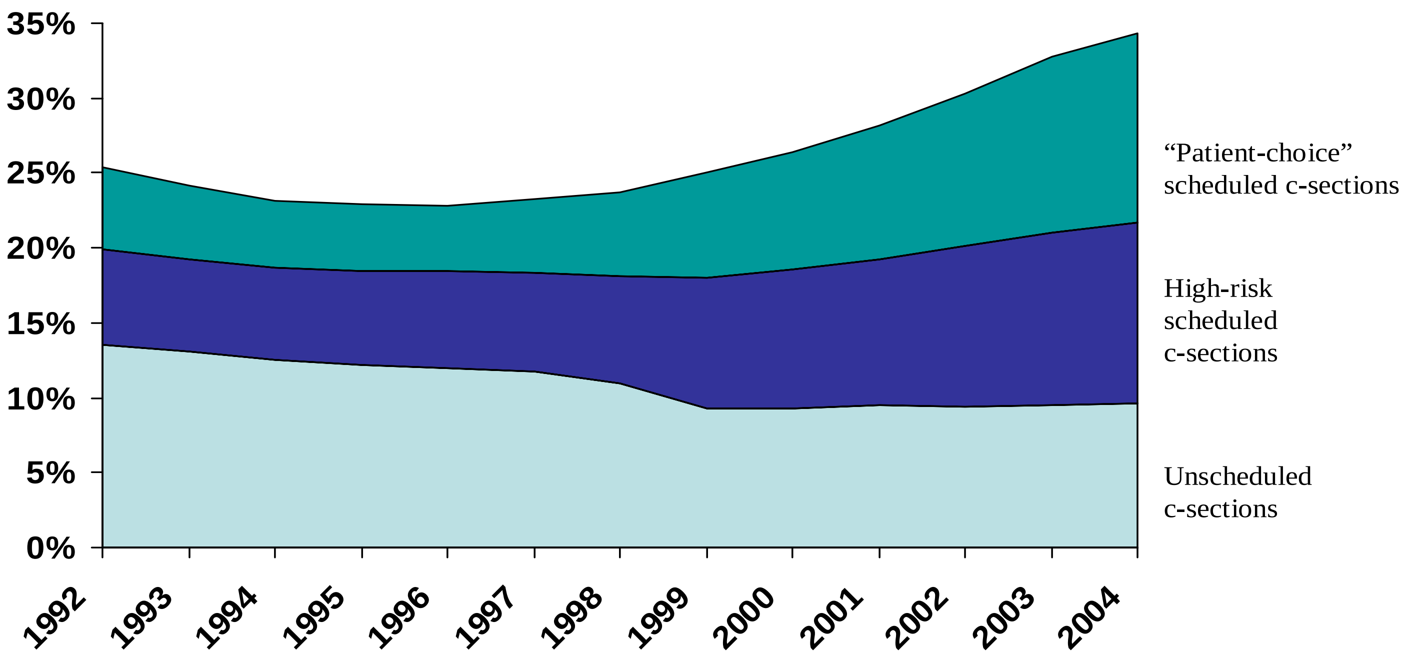
Notes: Ordinary least squares results for incremental length of stay due to major complications, controlling for group practice fixed effects, state, year, and patient and physician characteristics listed in the text. To account for heteroskedasticity, each observation is weighted by the inverse of the standard error of the dependent variable, which is the prediction from an auxiliary regression.
Robust standard errors in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Extent of Matching and Mismatching for Low-Risk Patients on Weekdays, by Health Condition and Practice Type

Patient Health Condition	Overall	Practice Type: Solo & Homogeneous Group	Practice Type: Differentiated Group	Practice Type: Differentiated Group - (Solo & Homogeneous Group)	P-value
Any of 12 risk factors					
Mismatched with Specialist	0.20	0.18	0.23	0.04	0.000
Matched with Avoider	0.34	0.37	0.30	-0.07	0.000
Preterm gestation					
Mismatched with Specialist	0.17	0.14	0.24	0.10	0.000
Matched with Avoider	0.32	0.32	0.33	0.01	0.000
Any hypertension					
Mismatched with Specialist	0.21	0.16	0.29	0.13	0.000
Matched with Avoider	0.29	0.29	0.27	-0.02	0.000
Malpositioned fetus					
Mismatched with Specialist	0.15	0.10	0.25	0.15	0.000
Matched with Avoider	0.20	0.19	0.23	0.04	0.000
Macrosomia					
Mismatched with Specialist	0.18	0.15	0.23	0.08	0.000
Matched with Avoider	0.29	0.32	0.22	-0.10	0.000
Maternal soft tissue disorder					
Mismatched with Specialist	0.18	0.17	0.20	0.04	0.000
Matched with Avoider	0.30	0.29	0.31	0.02	0.000
Pregnancy-related severe hypertension					
Mismatched with Specialist	0.13	0.10	0.19	0.10	0.000
Matched with Avoider	0.22	0.21	0.24	0.04	0.000
Antepartum bleeding					
Mismatched with Specialist	0.10	0.08	0.14	0.07	0.000
Matched with Avoider	0.12	0.09	0.20	0.11	0.000
Multiples (twins)					
Mismatched with Specialist	0.12	0.07	0.22	0.15	0.000
Matched with Avoider	0.17	0.13	0.26	0.13	0.000
Herpes					
Mismatched with Specialist	0.16	0.12	0.23	0.11	0.000
Matched with Avoider	0.21	0.20	0.23	0.03	0.000
Unengaged fetal head					
Mismatched with Specialist	0.16	0.17	0.15	-0.02	0.000
Matched with Avoider	0.29	0.23	0.38	0.15	0.000
Uterine scarring					
Mismatched with Specialist	0.08	0.08	0.09	0.01	0.000
Matched with Avoider	0.30	0.28	0.33	0.05	0.000
Congenital anomaly or chromosomal abnormality					
Mismatched with Specialist	0.09	0.07	0.12	0.06	0.000
Matched with Avoider	0.28	0.25	0.35	0.10	0.000
Unweighted average					
Mismatched with Specialist	0.14	0.12	0.20	0.08	
Matched with Avoider	0.25	0.23	0.28	0.05	

Figure 1

Percentage of C-section Deliveries in Florida by Type, 1992-2004

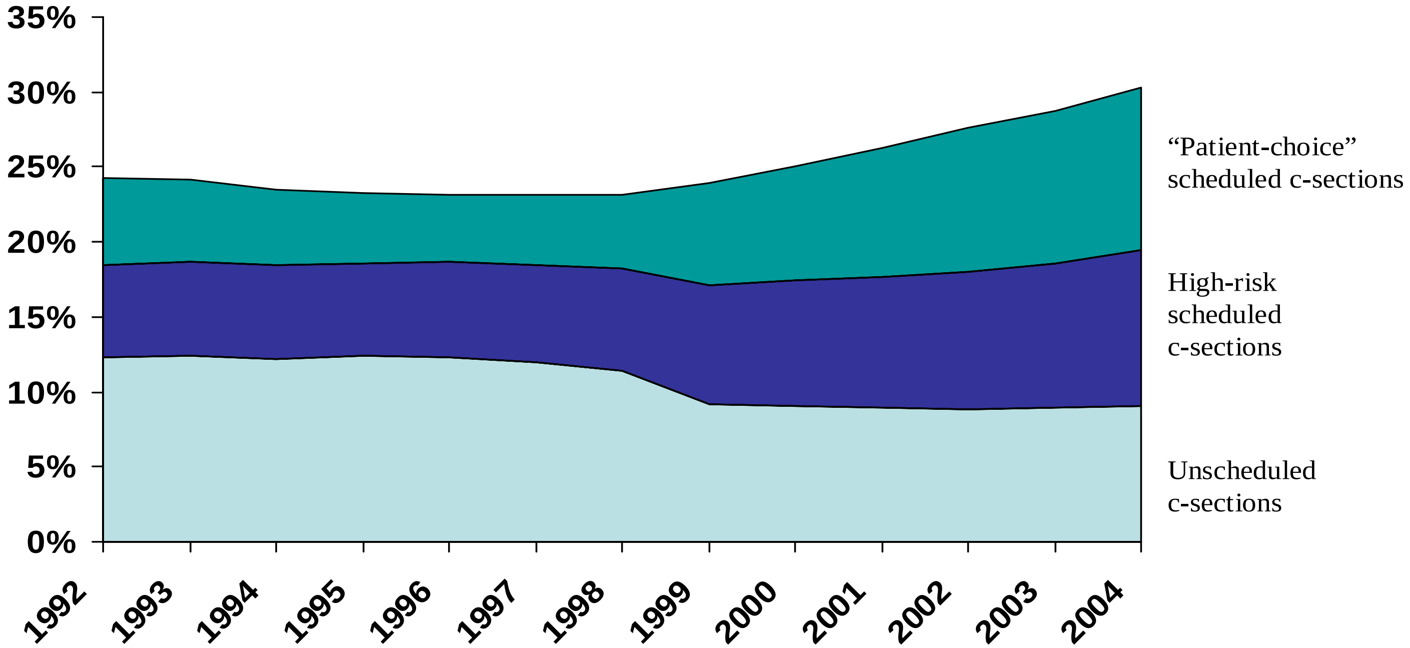


% of deliveries where woman labors	1992	1997	1999	2004
	88.2%	**88.5%**	**82.9%**	**75.2%**

Note: a delivery during which a woman does not labor is classified as a scheduled c-section. If a woman has any of the first 12 health conditions listed in Table 1, she is classified as "high risk."

Figure 2

Percentage of C-section Deliveries in New York by Type, 1992-2004



% of deliveries where woman labors	88.0%	88.8%	83.9%	78.7%

Note: a delivery during which a woman does not labor is classified as a scheduled c-section. If a woman has any of the first 12 health conditions listed in Table 1, she is classified as "high risk."

Table 6: Matching on Physician Style

Panel A. Group-level ratio of the weekday to weekend intra-group variances in physicians' risk-adjusted c-section rates (N=152 groups)

	Unadjusted	Bayesian Shrinkage
Groups with $\sigma^d_g > \sigma_g$	77	75
Groups with $\sigma^d_g < \sigma_g$	75	77
P value from Wilcoxon Signed-Rank Test	0.47	0.60
Median σ^d_g/σ_g	1.00	1.00
Bootstrapped P-value	0.84	0.97

Panel B: Correlation Coefficients to Test Mechanism of Sorting on Patient Preferences (N=407 physicians)

Overall	$Q^d_{js} - Q^d_{gs}$		
	Total	High-risk	Patient-choice
$\beta_{js} - \beta_{gs}$	0.136	0.120	0.137
P-value	[0.006]	[0.015]	[0.006]

Notes: All analyses restricted to groups that randomly pair patients and physicians on weekends, as measured by patient's predicted cesarean-section propensity.

Table 7: Matching on Physician Skill

Panel A: Matching based on a physician's absolute advantage for cesarean sections, relative to the group average (N=407 physicians)

Correlation coefficient	$(\xi_{jc} - \xi_{gc})$
Matching on observed clinical characteristics	
$\hat{C}^{d}_{js} - \hat{C}^{d}_{gs}$	-0.091
P-value from 1-sided test	[0.033]
Matching on unobserved preferences	
$\beta^{d}_{js} - \beta^{d}_{gs}$	-0.093
P-value from 1-sided test	[0.030]

Panel B: Matching based on a physician's comparative advantage for cesarean sections, relative to the group average (N=407 physicians)

Correlation coefficient	$\left(\frac{\xi_{jc}}{\xi_{jv}} - \frac{\xi_{gc}}{\xi_{gv}}\right)$
$\hat{C}^{d}_{js} - \hat{C}^{d}_{gs}$	-0.057
P-value from 1-sided test	[0.125]
$\beta^{d}_{js} - \beta^{d}_{gs}$	-0.073
P-value from 1-sided test	[0.071]

Notes: All analyses restricted to physicians in groups that randomly pair patients and physicians on weekends, as measured by patient's predicted cesarean-section propensity.

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

	Florida				New York			
	Vaginal Delivery		C-section		Vaginal Delivery		C-section	
Major Complication	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Secondary thrombocytopenia	0.489	0.147	0.626	0.230	0.162	0.092	0.265	0.161
Acute edema of lung	2.808	0.183	2.736	0.200	3.206	0.215	2.062	0.222
Urinary complications	0.419	0.194	0.723	0.285	0.992	0.221	1.374	0.387
Puncture or laceration by an instrument	0.767	0.237	0.581	0.168	0.766	0.260	0.625	0.177
Respiratory failure	4.856	0.299	3.592	0.233	2.888	0.326	3.358	0.289
3rd degree perineal laceration	0.153	0.013			0.131	0.010		
4th degree perineal laceration	0.228	0.019			0.168	0.017		
Laceration of the cervix	0.434	0.048			0.402	0.035		
High vaginal laceration	0.131	0.019			0.103	0.019		
Injury to pelvic organs	-0.031	0.015			-0.023	0.014		
3rd stage post-partum hemorrhage	0.513	0.044	1.486	0.163	0.638	0.035	0.871	0.125
Post-partum hemorrhage w/in 24 hours	0.331	0.017	0.313	0.061	0.301	0.015	0.344	0.056
Delayed post-partum hemorrhaging	0.663	0.042	0.872	0.272	0.519	0.039	1.025	0.219
Coagulation defects	0.682	0.065	0.835	0.103	0.705	0.060	0.739	0.098
Anesthesia complication	0.989	0.121	0.613	0.173	0.557	0.080	0.356	0.126
Major puerperal infection	1.752	0.060	1.745	0.046	2.093	0.042	1.612	0.033
Complication of OB surgical wound	0.839	0.085	1.060	0.068	0.719	0.075	1.166	0.051
Hemorrhage or hematoma	0.945	0.247	0.340	0.197	0.046	0.353	0.393	0.214
Other complications	1.621	0.119	1.140	0.183	1.397	0.094	1.514	0.160
Acute anemia due to loss of blood			0.299	0.035			0.308	0.042
Respiratory failure			3.197	0.429			2.820	0.408
Cardiac complications			1.042	0.385			0.861	0.256
Respiratory complications			1.152	0.207			1.108	0.268
Gastrointestinal complications			1.393	0.140			1.576	0.148
Anemia following the delivery			0.111	0.034			0.347	0.027
Retained placenta			0.219	0.184			0.545	0.153
Other OB wound surgical complication			1.026	0.082			1.119	0.073
Disruption of c-section wound			0.893	0.117			1.210	0.086
Postoperative infection			2.471	0.280			2.644	0.368
Other specified complications			1.039	0.163			1.286	0.227
Any Maternal Labor Dummy			0.319	0.012			0.355	0.010
Constant	2.487	0.017	3.720	0.060	2.276	0.020	3.892	0.056
N Deliveries	142,526		44,185		187,211		56,489	
R^2	0.088		0.151		0.057		0.198	

Notes: Models include hospital fixed effects. Standard errors are unadjusted.

Appendix B: OLS Coefficient Estimates of the Likelihood of Receiving a C-section in Florida and New York, 1999-2004

	Florida				New York			
	Weekend Laboring Deliveries		All Weekday Deliveries		Weekend Laboring Deliveries		All Weekday Deliveries	
Regressor	Coefficient	Std Err	Coefficient	Std Err	Coefficient	Std Err	Coefficient	Std Err
Patient Age	-0.005 ***	0.001	-0.007 ***	0.001	-0.003 ***	0.001	-0.004 ***	0.001
Patient Age Squared	0.000 ***	0.000	0.000 ***	0.000	0.000 ***	0.000	0.000 ***	0.000
Patient Age is Missing	-0.046	0.070	-0.048	0.045	0.021	0.069	0.084 **	0.041
Patient is Minority	0.004 **	0.002	0.003 ***	0.001	0.007 ***	0.002	0.009 ***	0.001
Patient Minority Status is Missing	-0.002	0.006	-0.001	0.003	-0.001	0.004	0.003	0.002
Medicaid or Uninsured	-0.013 ***	0.002	-0.019 ***	0.001	-0.020 ***	0.002	-0.026 ***	0.001
Previous C-Section Comorbidity	0.130 ***	0.004	0.634 ***	0.001	0.136 ***	0.003	0.601 ***	0.001
Malposition of Fetus Comorbidity	0.287 ***	0.004	0.494 ***	0.002	0.214 ***	0.004	0.487 ***	0.002
Antepartum Bleeding Comorbidity	0.087 ***	0.008	0.327 ***	0.003	0.076 ***	0.008	0.350 ***	0.004
Herpes Cormorbidity	0.010	0.007	0.207 ***	0.003	0.005	0.008	0.211 ***	0.005
Pregnancy-related Severe Hypertension Comorbiditity	0.176 ***	0.010	0.330 ***	0.004	0.188 ***	0.012	0.351 ***	0.005
Uterine Scar Unrelated to C-section Comorbidity	0.240 ***	0.035	0.465 ***	0.009	0.208 ***	0.036	0.490 ***	0.010
Multiple Gestation Comorbidity	0.105 ***	0.011	0.173 ***	0.004	0.072 ***	0.008	0.153 ***	0.004
Macrosomia Comorbidity	0.217 ***	0.005	0.283 ***	0.002	0.189 ***	0.007	0.273 ***	0.003
Unengaged Fetal Head Comorbidity	0.795 ***	0.007	0.675 ***	0.004	0.809 ***	0.006	0.695 ***	0.004
Maternal Soft Tissue Disorder Comorbidity	0.139 ***	0.006	0.178 ***	0.003	0.186 ***	0.006	0.205 ***	0.003
Other Types of Hypertension Comorbidity	0.093 ***	0.003	0.121 ***	0.002	0.079 ***	0.004	0.118 ***	0.002
Preterm Gestation Comorbidity	-0.058 ***	0.003	-0.012 ***	0.002	-0.045 ***	0.003	-0.012 ***	0.002
Congenital Fetal CNS Anomaly or Chromosomal Abnormality Comorbidity	0.019	0.028	0.152 ***	0.013	-0.011	0.030	0.210 ***	0.015
Cerebral Hemorrhage Comorbidity	0.368 *	0.200	0.172 ***	0.060	-0.118	0.186	0.304 ***	0.070
Asthma Comorbidity	0.025 ***	0.006	0.036 ***	0.003	0.031 ***	0.005	0.036 ***	0.003
Maternal Renal Abnormality Comorbidity	0.007	0.024	0.054 ***	0.011	0.060 **	0.025	0.062 ***	0.013
Maternal Liver Abnormality Comorbidity	0.032	0.026	0.044 ***	0.014	-0.032	0.029	0.006	0.018
Diabetes or Abnormal Glucose Tolerance Comorbidity	-0.005	0.008	0.036 ***	0.004	0.015	0.011	0.079 ***	0.007
Maternal Thyroid Abnormality Comorbidity	0.008	0.008	0.024 ***	0.004	0.007	0.006	0.015 ***	0.003
Maternal Substance Abuse	-0.006	0.019	0.003	0.012	-0.018	0.011	-0.024 ***	0.008
Mental Disorder Comorbidity	-0.015 ***	0.004	0.000	0.003	-0.002	0.005	0.013 ***	0.003
Maternal Congenital and Other Heart Disease Comorbidity	0.017 **	0.007	0.022 ***	0.004	0.007	0.006	0.019 ***	0.004
Isoimmunization Comorbidity	-0.011 *	0.006	-0.006 *	0.003	-0.005	0.005	-0.001	0.003
Intrauterine Fetal Demise Comorbidity	-0.109 ***	0.042	-0.275 ***	0.025	-0.056	0.050	-0.163 ***	0.026
Intrauterine Growth Restriction Comorbidity	0.000	0.009	0.107 ***	0.003	0.020 **	0.010	0.109 ***	0.004
Oligohydramnios Comorbidity	0.146 ***	0.012	0.143 ***	0.005	0.133 ***	0.012	0.125 ***	0.006
Polyhydramnios Comorbidity	0.075 ***	0.007	0.122 ***	0.003	0.075 ***	0.005	0.094 ***	0.002
Ruptured Membrane > 24 hrs Comorbidity	0.082 ***	0.006	0.046 ***	0.004	0.091 ***	0.005	0.038 ***	0.003

Appendix B (continued): OLS Coefficient Estimates of the Likelihood of Receiving a C-section in Florida and New York, 1999-2004

	Florida				New York			
	Weekend Laboring Deliveries		All Weekday Deliveries		Weekend Laboring Deliveries		All Weekday Deliveries	
Regressor	Coefficient	Std Err	Coefficient	Std Err	Coefficient	Std Err	Coefficient	Std Err
Chorioamnionitis Comorbidity	0.218 ***	0.005	0.209 ***	0.003	0.271 ***	0.005	0.262 ***	0.004
Other Maternal Infection Comorbidity	0.111 ***	0.006	0.078 ***	0.004	0.100 ***	0.006	0.077 ***	0.005
Uterine Rupture Comorbidity	0.469 ***	0.045	0.197 ***	0.016	0.466 ***	0.034	0.229 ***	0.017
Maternal Hypotension or Obstetrical Shock Comorbidity	-0.009	0.024	0.045 ***	0.012	0.032	0.023	0.052 ***	0.014
Pulmonary Embolism Comorbidity	0.128	0.142	0.332 ***	0.066	0.383 ***	0.108	0.303 ***	0.051
Year = 2000	0.002	0.002	0.014 ***	0.001	0.006 **	0.002	0.011 ***	0.002
Year = 2001	0.012 ***	0.003	0.032 ***	0.001	0.001	0.002	0.022 ***	0.002
Year = 2002	0.017 ***	0.003	0.050 ***	0.002	0.009 ***	0.002	0.032 ***	0.002
Year = 2003	0.019 ***	0.003	0.069 ***	0.002	0.010 ***	0.002	0.041 ***	0.002
Year = 2004	0.028 ***	0.003	0.082 ***	0.002	0.014 ***	0.002	0.056 ***	0.002
Intercept	0.102 ***	0.031	0.167 ***	0.017	0.062 **	0.027	0.174 ***	0.035
N Deliveries	150,124		708,692		162,571		618,769	
R-squared	0.17		0.43		0.18		0.43	

Notes: The dependent variable is an indicator for whether a delivery was by c-section. All models include physician fixed effects. Indicators are also included for whether a patient's race is missing and whether her age is missing.