NBER WORKING PAPER SERIES

THE PRODUCTION OF AND MARKET FOR NEW PHYSICIANS' SKILL

Andrew J. Epstein
Sean Nicholson
David A. Asch

Working Paper 18678
http://www.nber.org/papers/w18678

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

This paper benefited greatly from input from Jeph Herrin, Jason Hockenberry, David Howard, Suhui Li, Sindhu Srinivas, and comments from participants in the 2012 American Society for Health Economists conference, the 2012 Annual Health Economics Conference, and the 2012 Southeastern Health Economics Study Group meeting, the Penn Health Economics Workshop, and other seminars at the University of Pennsylvania. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Production of and Market for New Physicians' Skill
Andrew J. Epstein, Sean Nicholson, and David A. Asch
NBER Working Paper No. 18678
January 2013
JEL No. D83,I11,J24,L15

ABSTRACT

Our understanding of the determinants of physician skill and the extent to which skill is valued in the marketplace is superficial. Using a large, detailed panel of new obstetricians, we find that, even though physicians' maternal complication rates improve steadily with years of practice, initial skill (as measured by performance in a physician's first year of practice) explains most of the variation in physician performance over time. At the same time, we find that the trajectories of new physicians' delivery volume develop in a way partially consistent with Bayesian learning about physician quality. In particular, as physicians gain experience, their volume becomes increasingly sensitive to the information in their accumulated prior.

Andrew J. Epstein
Division of General Internal Medicine
University of Pennsylvania School of Medicine
423 Guardian Dr
1001 Blockley Hall
Philadelphia, PA 19104
eandrew@mail.med.upenn.edu

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

David A. Asch
Perelman School of Medicine
423 Guardian Drive
Philadelphia, PA 19104
asch@wharton.upenn.edu

1. Introduction

Although physicians have long been recognized for the essential role they play in the delivery of medical care, our current understanding of the formation and trajectory of their human capital development is superficial. This is captured in the literature on the "volume-outcomes relationship," which Luft (1980) described over 30 years ago as consisting of two simultaneous processes: learning by doing (in which physician skill improves with repetition) and selective referral (in which the market rewards better physicians with more patients). Owing to a lack of available data and endogeneity between volume and outcomes, prior work has been unable to reach firm conclusions about the determinants of physician skill and the extent to which skill is valued in the marketplace.

The need for a better understanding of physician skill is increasing as medical care systems transition toward an emphasis on provider quality and value. A number of strategies have been developed to motivate providers to improve the quality of their care, including public reporting and pay for performance. It would seem difficult to evaluate the success of these efforts or even gauge their potential for improvement without first appreciating how provider performance naturally evolves in their absence. Similarly, evaluation of efforts that are premised on steering consumers to better-performing providers, such as public reporting, depends on comparison to a baseline of consumers' ability to observe provider quality without additional guidance. There is scant evidence on the performance trajectories of individual physicians over time, however, and on consumers' ability to observe them.

We analyze these issues in a large sample of new obstetricians, whom we observe immediately after residency completion. Obstetrics is an attractive setting for a number of reasons. First, labor and delivery is one of the most common inpatient medical procedures in the

United States; in 2009, over 4 million women gave birth in hospitals.[1] As the number of deliveries per physician is large (120 per year on average in our sample), it is possible to measure physician performance relatively precisely. Second, delivery complications are commonplace. Approximately 11% of inpatient deliveries in 2006 resulted in a maternal complication (Srinivas et al., 2010), and there is significant variation across physicians (Asch et al., 2009). Thus, choice of provider for delivery matters for patient outcomes, which could be improved by better matching between patients and providers. Third, given the extended duration of pregnancy, patients have ample time and motivation to research provider choices and then choose the provider that best suits their needs (although the availability and use of external physician report cards in obstetrics have been minimal). Fourth, concerns about physician-induced demand of child births are unimportant (Dranove and Wehner, 1994), and the number of deliveries in the aggregate is presumably unrelated to the distribution of skill across practicing obstetricians. For these reasons, obstetrics offers a reasonably uncontaminated venue in which to study physician skill trajectories and as such represents an upper bound on unassisted market learning about physician skill.

We seek to answer three questions here. First, what is the trajectory of performance among new physicians? Second, which factors explain variation in performance across physicians over time? Specifically, how important are initial skill, economies of scale (i.e., contemporaneous volume), learning by doing (i.e., cumulative volume), and years of experience? Third, do physicians with better outcomes subsequently perform more deliveries? If so, is their volume more or less responsive to recent performance than to cumulative performance?

The paper proceeds as follows. In Section 2, we describe our data and methodology for constructing annual measures of physician skill. Section 3 contains an analysis of the

[1] http://www.cdc.gov/nchs/data/nhds/7femalesdelivery/2009fem7_numberpercent

determinants of physician performance over time, while in Section 4 we examine the market response to physician performance. Section 5 concludes.

2. Data and measurement of physician skill

2.1 Data and sample

We used data from Florida and New York all-payer hospital discharge databases for 1992 through 2010, covering all deliveries at all nonfederal acute care hospitals. These states were selected because their data contain physician identifiers in addition to information typically found in discharge databases, such as patient demographics, and diagnosis and procedure codes. Cesarean deliveries were identified with an *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) procedure code of 74 in any procedure field. Vaginal deliveries were identified with ICD-9-CM diagnosis codes of 650 or 640.0x through 676.9x (where x is 1 or 2) in the principal diagnosis field and no indication of a cesarean delivery. The discharge data were augmented with physician information from the American Medical Association's Physician Masterfile, including demographics, specialty and medical school and residency program locations and graduation years. We also used the American Medical Association's FREIDA Online graduate medical education website to identify hospitals that sponsored OB residency training programs or that hosted rotations.

There were 8,500,303 deliveries in the 1992-2010 hospital discharge data, of which 7,337,250 deliveries (86.3%) were performed by a physician with a valid state license number who completed an OB residency. Because physicians typically graduate from residency programs around July 1, we measured time according to an academic year calendar (starting July 1 and ending June 30); there were 6,968,506 deliveries (82.0%) performed in the 18 years between July 1, 1992 and June 30, 2010. As our focus is on new obstetricians, we identified 1,864 physicians who completed OB residency between 1992 and 2009 and appeared in the hospital discharge data in the same year (1,260 in New York and 604 in Florida); they collectively performed 2,004,903 deliveries over 15,673 physician-years. Sample descriptive statistics are shown in Table 1.

2.2 Measuring physician skill

Following our prior work (Asch, et al., 2009; Epstein, Ketcham and Nicholson, 2010; Srinivas, et al., 2010), we quantify skill based on physicians' annual rate of maternal complications of infection, hemorrhage, severe laceration, and other major operative and thrombotic complications. For vaginal deliveries we created a dichotomous composite variable that includes hemorrhage, severe laceration, infection and thrombotic complication. For cesarean deliveries we created a dichotomous composite variable that includes infection, hemorrhage, and other major operative and thrombotic complications. In addition to the two mode-specific complication rates, we constructed an overall complication rate that combined vaginal and cesarean deliveries. Maternal complications of these types are not uncommon; the unadjusted rate for any major maternal complication among deliveries in our sample was 15.0% in 1992 and 11.7% in 2010.

There are two analytic concerns related to measuring physician performance. First, maternal complications are determined in part by patient characteristics, and, second, there has been a secular trend in maternal complications (Srinivas, et al., 2010). Because patients are nested within physicians, we deal with these issues in a two-stage regression framework (Angrist and Pischke, 2009, pp. 313-315). In the first stage, we calculate a measure of each new obstetrician's annual performance based on the following linear probability model, which is estimated separately for each academic year during 1992-2009:

$$C_{ijt} = \alpha_t + \beta X_{ijt} + Y_{j_n t} + \varepsilon_{ijt} \quad (1)$$

where C_{ijt} equals one if patient i treated by physician j in year t experienced a maternal complication and is zero otherwise. New physicians are denoted by j_n. By including deliveries performed by both "ever-new" and "never-new"[2] obstetricians in the estimation sample, the coefficients on the new physicians' fixed effects ($\hat{Y}_{j_n t}$) are interpreted relative to the mean risk-adjusted complication rate of the established physicians[3] ($\hat{\alpha}_t$) in year t. We thus refer to $\hat{Y}_{j_n t}$ as physician j's normalized, risk-adjusted maternal complication rate for year t. The model controls for a vector of patient-level characteristics (X_{ijt}): age, race/ethnicity, having Medicaid or no insurance, weekend admission, and 34 maternal comorbidities.[4] Although we also control for delivery mode when the sample includes all deliveries, the results are not sensitive to doing so; the correlation between $\hat{Y}_{j_n t}$ with and without controlling for delivery mode is over 0.99.

Even though the outcome is dichotomous, we chose to estimate these models via OLS primarily because doing so avoids the incidental parameters problem resulting from including unconditional fixed effects in nonlinear models such as probit and logit. Johnson (2011) recently developed an approach to estimate $\hat{Y}_{j_n t}$ from nonlinear models in a correlated random effects framework, but the approach is computationally expensive. In comparison, OLS estimates are consistent, easily interpretable, and quickly estimated. Regardless, both approaches are designed to measure physician performance net of patient observables. Controlling for observable patient

[2] By "ever-new" we mean the set of those physicians whom we observe in our data immediately upon residency completion, while "never-new" physicians comprise the remainder.
[3] Note that by "established physicians" we mean those who completed OB residency prior to 1992 or practiced elsewhere before moving to Florida or New York.
[4] These include: prior cesarean delivery, fetal malpresentation, severe hypertension, multiple gestation, antepartum bleeding, herpes, macrosomia, unengaged head, maternal soft tissue disorder, preterm labor, congenital anomalies, oligohydramnios, and polyhydramnios.

characteristics does not affect the physician-specific maternal complication rate estimates much in the setting of obstetrics, because women delivering children are generally healthy. In our sample the correlations between $\widehat{Y}_{j_n t}$ with and without controlling for X_{ijt} are over 0.98. Neither approach accounts for possible sorting between patients and physicians based on unobservables, which have been shown to be relevant in obstetrics (Epstein, Ketcham and Nicholson, 2010). As a result, our measures of physician performance should be interpreted as reflecting "observed" physician performance and not necessarily "true" physician performance.

In the second stage, we pool the estimates of $\widehat{Y}_{j_n t}$ to create a panel dataset of 15,673 physician-year observations. We rescale t to represent the number of years since residency completion; e.g., t=0 indicates a physician's first year of practice. In analyses of the determinants of physician performance, we estimate linear models of the form:

$$\widehat{Y}_{j_n t} = \alpha + \beta Z_{j_n t} + \varepsilon_{j_n t} \tag{2}$$

where $Z_{j_n t}$ is a vector of possibly-time-varying physician-level covariates. Because of the heteroskedasticity resulting from unequal numbers of patients per physician-year in equation (1), equation (2) is estimated most efficiently by WLS, where the weights are the reciprocal of the variance of ε_{jt}. We use the formula for the weights developed by Borjas (1987), which is specific to situations in which the dependent variable in the second-stage model is a coefficient from a first-stage model. As WLS does not address the potential lack of independence among annual observations for a given physician or any residual heteroskedasticity, we also use robust, clustered standard errors.[5]

3. The production of physician skill

[5] Note that in models with physician fixed effects, robust standard errors are identical to standard errors that account for clustering of multiple physician-years per physician (Stock and Watson, 2008).

One of our main objectives is to identify the determinants of physician performance. Prior work on this topic has centered on the relationship between physician performance and experience, which may be measured in a variety of ways. One way is cumulative time spent in practice, via age or years of experience, for example. The standard model of human capital development is inverse-U-shaped and holds that human capital increases with age initially, before plateauing and eventually declining. Choudhry et al. (2005) reviewed the literature through 2004 and found weak evidence for a negative association between experience and physician performance, but noted that comparability across studies was limited and that most of the literature on physician age/experience focused on older surgeons and decrements in skill.

Another way to measure experience is in terms of numbers of cases treated or procedures performed. Underlying this approach is a learning-by-doing model, possibly with "forgetting" effects and scale effects (Huesch and Sakakibara, 2008). Although most volume-outcomes studies have analyzed data at the hospital level (Luft 1980), studies of physician volume and outcomes have become more common (e.g., Birkmeyer et al., 2003, and Janakiraman et al., 2011). These studies typically consider contemporaneous volume only (which is a measure of scale and not necessarily experience), employ a cross-sectional design, and fail to consider possible endogeneity. The few well-designed longitudinal studies at the physician level of which we are aware (Bridgewater et al., 2004; Vickers et al., 2007; Huesch, 2009; Ramanarayan, 2008) are split in terms of finding evidence of learning curves and are based on comparatively small samples of physicians (n=15, 72, 57 and 313 respectively).

To explore the relationship between experience and performance in our data, we first plot the association between years of experience and normalized, risk-adjusted maternal complication rates. We then fit simple models of

(3) $$\hat{Y}_{jt} = \alpha + \delta t_j + \varepsilon_{jt} \text{ and}$$

(4) $$\hat{Y}_{jt} = \alpha + \delta t_j + \mu_j + \varepsilon_{jt},$$

where t_j is the number of years elapsed since new physician j completed OB residency and μ_j represents a vector of physician fixed effects. Thus, Eq. (4) is identified off of within-physician changes in performance over time. Years of experience and cumulative volume have a naturally high correlation. We chose to focus here on years of experience rather than cumulative volume because of possible endogeneity between volume and performance and the difficulty of identifying a credible instrument for volume. Nonetheless, there is potential endogeneity between years of experience and performance in the form of survivor bias; poorly-performing physicians might be more likely to stop performing deliveries as they practiced longer, making it appear that more experience leads to better outcomes when in fact having worse outcomes leads to less experience. To address this concern, we repeated the analyses on the subset of "stayer" physicians—those who performed deliveries at the end of the study period and hence did not exit the sample.[6]

Figure 1 shows trends in mean, annual, physician-level normalized risk-adjusted maternal complication rates separately for all, cesarean, and vaginal deliveries. As physicians gained experience, performance improved for all three outcomes for about the first decade of practice. The estimation results for Eqs. (3) and (4) also indicate that average physician performance improved steadily (Table 2). Based on the unadjusted, weighted linear regressions (Eq. [3]), outcomes for all, cesarean and vaginal deliveries each improved by 0.18 percentage points per year over the study period. In the physician fixed effects models (Eq. [4]), the effects were attenuated slightly for all and vaginal deliveries and attenuated substantially for cesarean

[6] Approaches like this one using an unbalanced panel of physician-years might still be biased if physicians' start of practice (i.e., year of graduation) varies systematically with unobservables, such as innate ability. Our results were robust in sub-analyses using balanced panels of physicians with at least 5, 10 and 15 years of experience.

deliveries. For all three delivery samples, the results from the subset of "stayer" physicians were similar to the results from the full sample, suggesting that survivor bias is not a problem here.

While existing work has tested the influences of experience, age, volume and learning curves on physician performance, virtually no attention has been paid to the possible role of physicians' aptitude—some physicians may be innately better at performing some tasks than other physicians. We can explore this in a crude way by examining the correlation between a physician's risk-adjusted complication rate in year t and t-k. If aptitude plays a role, we would expect the correlations to be positive and significant. As shown in Table 3, these correlations are positive and significant through k=14. They range between 0.42 and 0.49 for k=1 and decrease steadily as the time lag increases (and the number of observations in our sample drops), suggesting that aptitude plays a role in extended performance.

Unlike years of experience and volume, we cannot measure innate aptitude directly. We therefore use physicians' normalized, risk-adjusted maternal complication rates in their first year of practice (i.e., t=0), which we refer to here as initial skill, as a proxy for innate ability. Figure 2 compares the performance-by-experience curves (starting from the second year of practice) for physicians in the first and fourth quartiles of initial skill. To test whether the differences between the first and fourth quartiles are significant, we calculated their difference and 95% confidence interval by year of experience using our WLS framework (Figure 3). For all, cesarean and vaginal deliveries, physicians who were in the best quartile of initial skill continued to have significantly better performance than their peers in the worst quartile of initial skill through at least the 15th year of practice. In other words, even as absolute performance steadily improves with years of experience, relative performance appears persistent.[7]

[7] As an alternative, we calculated the yearly proportion of physicians initially in the best or worst quartile who subsequently remained in that quartile. If assignment to performance quartile were random, we would expect 25%

One potential concern is regression to the mean. Intuitively it looks as though there is regression to the mean; performance between the two quartiles converges as physicians advance in their careers. In fact, regression to the mean is a feature intrinsic to any situation in which one follows subsequent performance conditional on initial performance (Nesselroade, Stigler and Baltes, 1980; Smith and Smith, 2005). Under the assumption that the relationship between latent physician aptitude and realized performance is time-invariant, we could quantify the magnitude of the regression to the mean as one minus the correlation coefficient. Data from Table 3 suggest that the size of year-to-year regression to the mean would be just over 50%. But, the aptitude-performance relationship almost surely changes as physicians gain experience, and our interest here is in determining how much it changes. That is, the point of our exercise is not to establish whether or not regression to the mean is present (it is by construction) or to measure its magnitude (which is sizeable), but to investigate how long it takes for the performance differences identified in the first year to dissipate (at least 15 years).

Given this evidence that experience and initial skill both affect physician performance, we would like to develop a production function for physician skill. The existing work on production functions for physician practices (e.g., Reinhardt, 1972; Thurston and Libby, 2002) focuses exclusively on quantities of services provided (e.g., numbers of office visits) as a function of physician labor and other inputs. In contrast, we seek to model physicians' average quality level conditional on their quantity of deliveries performed in a given year. We are interested in quantifying the relative contributions of four potential determinants of physician performance: (1) aptitude (i.e., initial skill), (2) scale (i.e., contemporaneous volume), (3)

---

to remain in the same quartile between any pair of years. For the all deliveries outcome measure, the proportion declined from 51% in the second year of practice to 37% in the sixth year of practice before stabilizing between 32% and 39% through the 18th year of practice. Patterns for the cesarean and vaginal delivery measures were similar.

learning by doing (i.e., cumulative volume), and (4) years of experience. However, as noted, cumulative volume and years of experience are highly correlated. Thus we either consider their joint contribution to performance, or control for it implicitly by looking only at physicians with the same amount of experience. Our simple linear model of performance is

(5) $$\hat{Y}_{jt} = \alpha + \beta\hat{Y}_{j0} + \gamma Vol_{j(t-1)} + \delta \sum_{s=0}^{t-2} Vol_{js} + \zeta t_j + \varepsilon_{jt}$$

where $\hat{Y}_{j0}$ is initial skill, $Vol_{j(t-1)}$ is lagged contemporaneous volume, and $\sum_{s=0}^{t-2} Vol_{js}$ is lagged cumulative volume for t > 0. We lag contemporaneous volume one year to help attenuate reverse causality concerns, and compute cumulative volume only through year t-2 to prevent a mechanical correlation with contemporaneous volume. Of course, volume and outcomes may still be endogeneous. Assuming more volume accrues to better performers, as is consistent with theory and prior literature (e.g., Luft, 1980), our estimate of the contribution of volume to performance is an upper bound.

We use Theil's (1972) incremental R^2 method to decompose the total explained variation into components uniquely attributable to each factor. As described above, these models will be heteroskedastic because the dependent variables are physician-specific fixed effect coefficients estimated in the first-stage, patient-level regressions (Eq. [1]). We nevertheless estimate these models with OLS, because WLS transforms the regressand and regressors and so the R^2 obtained from WLS is not applicable to the study question (Wooldridge, 2006, p. 286). To check this, we estimate one set of models on a sample limited to physician-year observations in which physicians performed at least 30 deliveries of the relevant type.

The incremental R^2 method consists of two steps. The first step is to calculate how much of the variation in the dependent variable can be explained in total by all of the available explanatory variables. We thus estimate Eq. (5). The R^2 value from this regression is reported in

Table 4 in the "Total R^2" rows. The second step is to calculate how much of the variation in the dependent variable can be explained by the other explanatory variables except for the one(s) of interest. We thus re-estimate Eq. (5), omitting each of the explanatory variables one at a time. The unique contribution of each explanatory variable to the variation in physician performance is calculated as the simple arithmetic difference between the R^2 from the full model and the R^2 from the model omitting that explanatory variable. These are displayed in Table 4 in the same rows next to the particular explanatory variable. We do not report the overall R^2 values from these regressions, but they can be calculated straightforwardly from Table 4.

Table 4 shows the results of the decomposition exercise for a number of distinct samples of physicians. The first three columns are based on analyses of all years of physician experience except the first, which was omitted to avoid the reflection with our measure of initial skill. The first column includes all physician-year observations, the second includes physician-year observations where mode-specific delivery volume was at least 30 (as a way to check the effects of heteroskedasticity without resorting to WLS), and the third is like the first but includes observations from "stayer" physicians only. The second three columns are based on analyses in which physician experience is held fixed, at 5 years, 10 years and 15 years of experience respectively.

The results are notably stable across samples. Initial skill contributes around 80% of the explained variation in annual physician performance for all, cesarean and vaginal deliveries in the "all years" samples (the first three columns).[8] In contrast, for all three delivery types, contemporaneous volume contributes less than 5%, and cumulative volume and years of

---

[8] We also estimated models in the "all years" samples that controlled only for physician fixed effects to gauge the amount of variation in $\hat{Y}_{jt}$ explained by time-invariant physician attributes all together. The R^2 values from these models were 48.16 for all deliveries, 42.81 for cesarean deliveries and 45.02 for vaginal deliveries—all notably higher than the Total R^2 values reported in Table 4, which were between 4 and 9.

experience together contribute no more than 10%. Initial skill continues to explain more of the variation in performance than the combination of contemporaneous and cumulative volume when holding years of experience fixed, even as the absolute amount of explained variation (and the number of physicians in the sample) declines with years of experience. After five years of experience (i.e., in the sixth year of practice), initial skill is responsible for 85% of the explained variation in performance among all deliveries, 96% among cesarean deliveries, and 75% among vaginal deliveries. Even after 15 years of experience, the contribution of initial skill remains at 38%, 69% and 56% respectively. Interestingly, the relative importance of contemporaneous volume increases as physicians practice longer, accounting for less than 2% of explained variation in the sixth year of practice to between 13% and 20% in the 16th year. This may be evidence of "forgetting" effects in the obstetrics learning curve or of delayed specialization by physicians. Lagged cumulative volume is responsible for no more than 5% of explained variation for all and cesarean deliveries in any single year. For vaginal deliveries, however, it contributes 11% after five years of experience before declining to 1% after 15 years of experience. Similarly, in the "all years" samples, cumulative volume and years of experience jointly contribute more to explained variation for vaginal deliveries than for cesarean deliveries.

To help put these findings in perspective, we can treat Eq. (5) as a causal model and compare the expected impact of physician attributes on complication rates. Suppose a hypothetical consumer could observe information about quartile of initial skill, cumulative volume and years of experience for a pair of prospective obstetricians.[9] Still taking the results of Eq. (5) as causal, we would expect that an obstetrician with 17 years of experience (and average cumulative volume among physicians with 17 years of experience) would have a 2.9 percentage point lower normalized, risk-adjusted complication rate for all deliveries compared with a peer

[9] We ignore contemporaneous volume here, as its estimated impact is minimal.

with only one year of experience (and average cumulative volume among physicians with one year of experience). In contrast, a physician in the best quartile of initial skill would have a 5.4 percentage point lower complication rate for all deliveries on average than an otherwise-identical peer in the worst quartile. In other words, a consumer could improve her expected outcome by 2.5 percentage points ($p<0.001$) by selecting an obstetrician based on quartile of initial skill instead of the combination of cumulative volume and years of experience.[10] Further, note that this hypothetical scenario was constructed in a way to favor information on cumulative volume and years of experience (by selecting the extreme ends of experience and by assuming there is no reverse causality between outcomes and volume) over initial skill. One important caveat, however, is that the size of the absolute advantage of selecting based on initial skill applies only to physicians with no more than 17 years of experience (i.e., physicians in our sample). The results in Table 4 suggest that initial skill is relatively more informative even 15 years into a physician's career, but its absolute predictive power declines with experience.

An important caveat here is that initial skill may be determined by more than just a physician's innate aptitude. Contextual factors, such as the resources of the hospital in which the physician works and the skill of complementary labor inputs, likely matter as well. This has been recognized in the labor literature on the impact of graduates' initial job placement on subsequent career trajectories (e.g., Oyer, 2006). A basic way to address this concern is to expand Eq. (5) to include fixed effects for the hospitals at which physicians practiced in their first years after residency training. Results from these expanded models, which are shown in Table 5, indicate that the hospitals at which obstetricians begin their careers have a sizeable downstream impact on their later performance. Given that most new obstetricians in our sample

[10] The results are similar for the cesarean delivery and vaginal delivery samples. The expected net benefit of choosing an obstetrician based on initial skill is a 3.3 percentage point lower complication rate for cesarean deliveries and a 3.0 percentage point lower complication rate for vaginal delivery ($p<0.001$ for each).

do not change hospitals during the study period, these initial hospital fixed effects are also capturing environmental factors in year t as well. At the same time, the relative importance of physicians' initial skill diminishes considerably. When controlling for initial hospital fixed effects, initial skill contributes between 2% and 11% of explained variation, still more than volume or experience, but less than the 58% to 76% from the initial hospital fixed effects.[11]

Taken as a whole, our analyses suggest that, while obstetricians' performance continues to improve with years of experience, their relative positions are fairly stable over time and are reflected in a composite measure of their performance in their first years of practice. This composite measure in turn reflects components that are specific to physicians and components that are specific to practice settings. Because the selection of physicians into practice settings is endogenous, we cannot easily disentangle the relative importance of each component. Nevertheless, the measure of physicians' initial skill retains considerable predictive power years later, implying that physician ability, regardless of its determinants, is preserved over time, at least during the first portion of a physician's career.

4. The market for physician skill

A small number of studies find that market share responds to provider quality. Early studies focused primarily on hospital quality (e.g., Luft et al., 1990), while more recent efforts also look at physicians. Consistent with the hospital studies, two studies on the selection of cardiac surgeons (Mukamel et al., 2004/2005; Epstein, 2010) found that patients were more

[11] We also estimated models that replaced the initial hospital fixed effects with "traditional" hospital fixed effects and physician j's initial hospital's mode-specific crude performance in their year prior to physician j's start. Physician initial skill and initial hospital (lagged) performance contributed roughly the same amount to explained variation in subsequent physician performance. Initial hospital performance is slightly more important in cesarean deliveries, while physician initial skill is more important for vaginal deliveries. Moreover, the correlation between the two is large (e.g., 0.50 for all deliveries).

likely to be treated by better-rated surgeons and less likely to be treated by worse-rated ones during periods when provider report cards were not available. These studies all cover narrow time periods and thus offer a limited view of the relationship between prior performance and current volume.

Two recent studies seek to overcome these limitations by using panels of physicians. Navathe and David (2009) explore the market volume response to physician performance in the prior year while controlling for physician fixed effects and other time-varying characteristics over a 15-year period. Johnson (2011) uses an initial cross-section of data to measure physician quality and looks at responses to the quality during the following six. An important feature of both is their inclusion of all physicians—both new and established—in their analytic samples. Under certain conditions, however, recent performance may not influence market participants' beliefs about the quality of established physicians. By limiting our study to new physicians and following them from residency completion, we hope to expand the understanding of how participants in the market form and then update their expectations of physician quality.

In the setting of obstetrics services, where patients have adequate time and incentive to search for providers, the aggregate market response will be driven by a combination of patient demand, peer and referring physician demand, and obstetrician supply. It is entirely possible that any group may not respond at all to obstetricians' performance, perhaps because it is not observable to them or because it is not trusted or deemed relevant. We posit that patients obtain their information primarily from word of mouth around the time of their pregnancies (Hoerger and Howard, 1995). Thus, we would expect the information signal to decay quickly, and if patients respond, it would be primarily in reaction to obstetricians' recent performance. In contrast, physicians (peer/referring and obstetricians themselves) could develop the capacity to track obstetricians over time as their careers unfold. This would allow for the accumulation of information about a physician's performance and would enable long-run learning. It is an empirical question as to whether physicians accumulate performance information and respond to it; alternatively, like patients, they might respond principally to obstetricians' recent performance if at all.[12]

Like Navathe and David (2009) and Johnson (2011), we specify a simple Bayesian model of learning about physician quality over time in which the expectation of quality is the weighted average of a recent quality signal and an accumulated prior belief about physician j's quality. That is,

(6) $$E(Q_{jt}) = Posterior_{jt} = w_{jt}Signal_{jt} + (1 - w_{jt})Prior_{jt}$$

where $E(Q_{jt})$ is the market's expectation of quality for physician j in time t. We further assume that $Signal_{jt} \sim Normal(\hat{Y}_{jt}, 1/Vol_{jt})$ for t ≥ 0. The market prior about a physician's quality is initially flat $\left(Prior_{j0} \sim Normal(0, \infty)\right)$ and is updated dynamically as $Prior_{jt} = Posterior_{j(t-1)}$ for t > 0. Thus, the market's expectation of physician quality becomes

(7) $$E(Q_{jt}) = \frac{Vol_t}{\sum_{k=0}^{t} Vol_k} Signal_{jt} + \frac{\sum_{k=0}^{t-1} Vol_k}{\sum_{k=0}^{t} Vol_k} Prior_{jt},$$

which simplifies to

(8) $$E(Q_{jt}) \approx \frac{1}{t} Signal_{jt} + \frac{t-1}{t} Prior_{jt}$$

when physician j's volume is constant over time. So, if there is learning in a Bayesian fashion, we would expect that the response to a physician's recent performance is strongest when the physician is at the start of his career (i.e., when t = 1, $E(Q_{jt}) = Signal_{jt}$). As the physician gains

[12] There are other stakeholders who might accumulate information about physician performance over time, such as patient groups or payers, although we suspect these are less likely mechanisms.

experience, however, market participants would increasingly rely on the information in the accumulated prior (i.e., as $t \rightarrow \infty$, $E(Q_{jt}) \rightarrow Prior_{jt}$) if they are Bayesian learners.

We then test the predictions from the learning model in the following four empirical specifications.

$$(9) \quad Vol_{jt} = \alpha + \beta_1 Doc_j + \beta_2 Hosp_j + \zeta_t + \gamma_0 Sig_{jt} + \gamma(Sig_{jt} \times \zeta_t) + \delta_0 Pri_{jt} + \delta(Pri_{jt} \times \zeta_t) + \eta_s + \varepsilon_{jt}$$

$$(10) \quad Vol_{jt} = \alpha + \beta_1 Doc_j + \beta_2 Hosp_j + \zeta_t + \gamma_0 Sig_{jt} + \gamma(Sig_{jt} \times \zeta_t) + \delta_0 Pri_{jt} + \delta(Pri_{jt} \times \zeta_t) + \theta_m + \varepsilon_{jt}$$

$$(11) \quad Vol_{jt} = \alpha + \beta_1 Doc_j + \quad 0 \quad + \zeta_t + \gamma_0 Sig_{jt} + \gamma(Sig_{jt} \times \zeta_t) + \delta_0 Pri_{jt} + \delta(Pri_{jt} \times \zeta_t) + \kappa_h + \varepsilon_{jt}$$

$$(12) \quad Vol_{jt} = \alpha + \quad 0 \quad + \zeta_t + \gamma_0 Sig_{jt} + \gamma(Sig_{jt} \times \zeta_t) + \delta_0 Pri_{jt} + \delta(Pri_{jt} \times \zeta_t) + \lambda_j + \varepsilon_{jt}$$

In these models, Vol_{jt} represents physician j's delivery volume in academic year t, Doc_j is a vector of time-invariant physician characteristics (i.e., sex, maternal fetal medicine specialist, international medical graduate status, whether the obstetrician practiced in multiple hospitals, and indicators for the year of OB residency completion), $Hosp_j$ is a vector of time-invariant characteristics of physician j's main hospital (i.e., the average number of annual deliveries and indicators for whether the hospital sponsored an OB residency program or whether the hospital hosted OB residency rotations), ζ_t is vector of indicators for the number of years since residency completion, Sig_{jt} is the signal, Pri_{jt} is the prior, η_s is an indicator for state, θ_m represents market fixed effects, κ_h represents hospital fixed effects, and λ_j represents physician fixed effects. Note that observations from t = 0 are dropped from the estimation sample because the signal is not defined in the first year of practice. When t = 1, the prior is set to 0 according to the assumption of a flat initial prior (and 0 is the normalized mean performance from established physicians in each calendar year).

Our principal interest is in the net effects of each of the signal and prior across the values of t. Because the available sample decreases as t grows, in the interactions with ζ_t we collapse categories of t into $t \in \{1, 2, 3, 4, 5-6, 7-11, 12-17\}$, where t = 1 is the omitted referent. Thus, to obtain the net effect of, for example, the signal in a given year, we sum γ_0 and the appropriate member of γ. These net effects based on OLS models with robust, clustered standard errors[13] are presented for all deliveries (Table 6), cesarean deliveries (Table 7) and vaginal deliveries (Table 8). To ease interpretation, we present the net effects of a one standard deviation increase in the prior or signal on annual delivery volume.

The estimated net effects of the information on recent performance contained in the signal on physician volume do not follow an obvious pattern. For all and vaginal deliveries, the signal effect is generally negative and largest in the third year of practice—at least for Eqs. 9, 10 and 11. The physician fixed effects models (Eq. 12) for all and vaginal deliveries and all four models for cesarean deliveries reveal no reliable relationships between the signal and current volume. At most, we could say that there is occasional evidence that delivery volume responds to physicians' recent performance.

In contrast, the market appears to respond to the performance information accumulated in the prior. For all, cesarean and vaginal deliveries, across nearly all model specifications, the association between the prior and volume is small after a physician completes residency and starts practice, and it grows nearly monotonically in the right direction (i.e., the net effects

[13] Note that we do not use WLS here, as the dependent variable is volume and not physician performance. Because volume is skewed, however, we also estimated quasi-maximum likelihood Poisson models. The patterns of coefficients from these models are the same, so we report only the OLS models.

become more negative, as expected given that the prior and signal are measured in terms of maternal complication rates, for which higher scores are worse). By the eighth year of practice, the response to the prior is nominally larger than the response to the signal. The one exception to this pattern is the physician fixed effects model (Eq. 12) for vaginal deliveries, in which the prior shows no consistent pattern over time.

So far, our analyses have shed no light on the extent to which the volume response to quality information is driven by consumers or suppliers (or both). To test whether consumers respond, we follow others in exploring whether the volume response varies by patients' insurance type (Fournier and McInnes, 2002; Dranove, Ramayanaran and Watanabe, 2012), specifically comparing mothers with Medicaid or no insurance to those with commercial insurance.[14] Our identifying assumptions here are that obstetricians cannot induce demand for delivery (Dranove and Wehner, 1994); that obstetricians do not face capacity constraints; and that, as obstetricians learn about their own relative performance, they do not reduce their own volume differentially by patient insurance status. Thus, finding that commercially-insured patients were more responsive would suggest that at least some of the response was consumer-based.

We test this empirically by re-estimating Eqs. 9-12 after limiting Vol_{jt} to the relevant group of patients based on their insurance type. The all delivery results for commercial patients are shown in Table 9 and for Medicaid/uninsured patients in Table 10. As with the overall results, while we would expect the effect of the signal to be large and negative for obstetricians early in their careers and then attenuate for obstetricians with more years of experience, there is no clear pattern between the quality signal and delivery volume for either patient group. If

[14] Fournier and McInnes (2002) and Dranove, Ramayanaran and Watanabe (2012) both use data from Florida which code HMOs and PPOs separately. We combine all commercial patients because the New York data do not indicate type of plan.

consumers learn in a Bayesian fashion, their response to the prior should be small for the newest physicians and grow more negative for more experienced physicians. The response pattern among commercial patients appears consistent with this; the magnitude of the effect of the prior grows from the 3rd year to the 6th-7th year period to the 8th-12th year period, and stays larges in the 13th-18th year period. The response to the prior among Medicaid and uninsured patients stays largely flat through the 8th-12th year period before becoming large and negative in the 13th-18th year period. Thus our results offer mild support for the hypothesis that commercially insured patients are more responsive to the physician performance information contained in the prior, but there is no substantive difference between the groups in their response to the signal.

We can study whether obstetricians are responding to their own performance information by examining whether they increase their volume of substitute procedures. If so, we would expect that the gynecology procedure volume of poorly-performing obstetricians would increase relatively more. Alternatively, if physicians with better gynecology procedure outcomes also attract more volume and obstetrics and gynecology procedure skill are positively related, then physicians who perform poorly at obstetrics procedures might lose more gynecology procedure volume too. We thus re-estimate Eqs. 9-12 where the outcome is measured as annual number of common inpatient gynecological procedures (oophorectomies and hysterectomies); the results are shown in Table 11. If obstetricians learn about their own performance in a Bayesian fashion and substitute toward gynecology, we would expect their response to the signal to be large and positive early in their careers and attenuate over time, while their response to the prior should be close to zero early on before growing large and positive with years of experience. However, the effect of the prior starts small but grows increasingly negative with years of experience (except for the physician fixed effects model, in which the response is close to zero throughout), suggesting if anything that physicians who establish themselves as performing better in obstetrics also attract more gynecological volume.

Another possibility is that physicians react to learning that their obstetrics performance is poor by exiting obstetrics entirely.[15] We adapted the specification in Eq. 9 to model the hazard of physician exit from obstetrics in year t (i.e., t is the first year with zero deliveries, and there are no deliveries thereafter) using a discrete-time implementation of a survival model based on conditional-log-log regression. Results, which are shown in Table 12, do not indicate that physicians were basing their exit decisions on revelations about their obstetrics performance as measured in the signal or prior. These stand in contrast to Johnson's (2011) findings that worse-performing cardiac specialists were more likely to exit practice and to change geographic markets.

5. Conclusions

Our empirical analyses indicate that, in the setting of obstetrics, average physician skill improves steadily in an almost linear fashion over the first decade of a new obstetrician's practice. At the same time, the importance of any experience-based learning curve is dominated by physicians' initial skill, which reflects some combination of innate ability and practice setting characteristics. Compared with physicians in the worst quartile of initial skill, those in the best quartile remain significantly better for more than a decade after. Moreover, initial skill has far more power in predicting physician performance than contemporaneous volume, cumulative volume and years of experience; it contributes around 80% of the explained variation in models using our full sample. Perhaps most importantly, consumers could substantially improve their

[15] We cannot determine from our data whether an exiting physician has moved another state or why the physician stopped performing deliveries.

expected reduction in complications by selecting prospective obstetrician candidates based on their initial skill rather than their cumulative delivery volume and years of experience if this information were known.

We find evidence partially consistent with Bayesian learning by the market, most likely involving peer and referring physicians. Response to the information contained in the accumulated prior is minimal at the start of a physician's career and increases steadily in the expected direction going forward. By the eighth year of practice, response to the prior is nominally but consistently larger than response to the signal of recent performance. Response to the signal is somewhat unstable, however. If learning were strictly Bayesian, we would expect the signal response to be largest for recently-graduated physicians and attenuated with experience. For all and vaginal deliveries, response to the signal is large in the third year of practice, but it is large and in the right direction in some later years as well. For cesarean deliveries, the response is in the right direction but the magnitudes follow no obvious pattern. These results may reflect a distinct, non-Bayesian process in which patients and/or other physicians respond to recent performance, or they may reflect coincidence.

There are a number of potential limitations to this work. One category is generalizability. We examine data on physicians from one specialty and two states. The patterns found here may not apply beyond obstetrics. Another category of limitations is measurement error. We use administrative data to identify maternal complications and attribute them, possibly inaccurately, to individual physicians. These data do not allow attribution of neonatal outcomes, which are also important, to obstetricians. These data are also incomplete in measuring potentially relevant patient characteristics that might be related to patient selection and sorting. More broadly, we do not observe details about physicians' practice environments, including group structure, call schedules and other factors that might relate to patient and treatment selection. Finally, we do not observe the physician job selection process, which might be endogenous with subsequent performance if, for example, physicians match to jobs for which they are well-suited.

References

Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, NJ.

Asch, D.A., Nicholson, S., Srinivas. S., Herrin. J., Epstein, A.J., 2009. Evaluating obstetrical residency programs using patient outcomes. Journal of the American Medical Association 302 (12), 1277-1283.

Birkmeyer, J.D., Stukel, T.A., Siewers, A.E., Goodney, P.P., Wennberg, D.E., Lucas, F.L., 2003. Surgeon volume and operative mortality in the United States. New England Journal of Medicine 349, 2117-2127.

Borjas, G.J.,1987. Self-selection and the earnings of immigrants. American Economic Review 77 (4), 531-553.

Bridgewater, B., Grayson, A.D., Au, J., Hasan, R., Dihmis, W.C., Munsch, C., Waterworth, P., 2004. Improving mortality of coronary surgery over first four years of independent practice: retrospective examination of prospectively collected data from 15 surgeons. British Medical Journal 329 (7463), 421.

Dranove, D., Ramanarayanan, S., Watanabe, Y., in press. Delivering bad news: market responses to negligence. Journal of Law and Economics.

Dranove, D., Wehner, P., 1994. Physician-induced demand for childbirths. Journal of Health Economics 13 (1), 61-73.

Epstein, A.J., 2010. Effects of report cards on referral patterns to cardiac surgeons. Journal of Health Economics 29, 718-731.

Epstein, A.J., Ketcham, J.D., Nicholson, S., 2010. Specialization and matching in professional services firms. RAND Journal of Economics 41 (4), 811-834.

Epstein, A.J., Nicholson, S.S., 2009. The formation and evolution of physician treatment styles: an application to cesarean sections. Journal of Health Economics 28 (6), 1126-1140.

Fournier, G.M., McInnes, M.M., 2002. The effects of managed care on medical referrals and the quality of specialty care. Journal of Industrial Economics 50 (4), 457-473.

Hoerger, T.J., Howard, L.Z., 1995. Search behavior and choice of physician in the market for prenatal care. Medical Care 33 (4), 332-349.

Huesch, M.D., 2009. Learning by doing, scale effects, or neither? Cardiac surgeons after residency. Health Services Research 44 (6), 1960-1982.

Huesch, M.D., Sakakibara, M., 2008. Forgetting the learning curve for a moment: how much performance is unrelated to own experience? Health Economics 18 (7), 855-862.

Janakiraman, V., Lazar, J., Joynt, K.E., Jha, A.K., 2011. Hospital volume, provider volume, and complications after childbirth in U.S. hospitals. Obstetrics & Gynecology 118 (3), 521-527.

Johnson, E.M., 2011. Ability, learning and the career path of cardiac specialists. Massachusetts Institute of Technology working paper.

Luft, H.S., 1980. The relation between surgical volume and mortality: an exploration of causal factors and alternative models. Medical Care 18 (9), 940-959.

Luft, H.S., Garnick, D.W., Mark, D.H., Peltzman, D.J., Phibbs, C.S., Lichtenberg, E., McPhee, S.J., 1990. Does quality influence choice of hospital? Journal of the American Medical Association 263 (21), 2899-2906.

Mukamel, D.B., Weimer, D.L., Zwanziger, J., Gorthy, S.H., Mushlin, A.L., 2004/2005. Quality report cards, selection of cardiac surgeons, and racial disparities: a study of the publication of the New York state cardiac surgery reports. Inquiry 41 (4), 435-446.

Navath, A., David, G., 2009. The formation of peer reputation among physicians and its effect on technology adoption. Journal of Human Capital 3 (4), 289-322.

Nesselroade, J.R., Stigler, S.M., Baltes, P.B., 1980. Regression toward the mean and the study of change. Psychological Bulletin 88 (3), 622-637.

Oyer P. 2006. Initial labor market conditions and long-term outcomes for economists. Journal of Economic Perspectives 20(3):143-160.

Ramanarayanan, S., 2008. Does practice make perfect: an empirical analysis of learning-by-doing in cardiac surgery. University of California, Los Angeles working paper.

Reinhardt, U., 1972. A production function for physician services. Review of Economics and Statistics 54 (1), 55-66.

Smith, G., Smith, J., 2005. Regression to the mean in average test scores. Educational Assessment 10 (4), 377-399.

Srinivas, S.K., Epstein, A.J., Nicholson, S., Herrin, J., Asch, D.A., 2010. Improvements in US maternal obstetrical outcomes from 1992 to 2006. Medical Care 48 (5), 487-493.

Stock, J.H., Watson, M.W., 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. Econometrica 76, 155-174.

Theil, H., 1971. Principles of Econometrics. John Wiley & Sons, Inc.: New York, NY.

Thurston, N.K., Libby, A.M., 2002. A production function for physician services revisited. Review of Economics and Statistics 84 (1), 184-191.

Vickers, A.J., Bianco, F.J., Serio, A.M., Eastham, J.A., Schrag, D., Klein, E.A., Reuther, A.M., Kattan, M.W., Pontes, J.E., Scardino, P.T., 2007. The surgical learning curve for prostate cancer control after radial prostatectomy. Journal of the National Cancer Institute 99, 1171-1177.

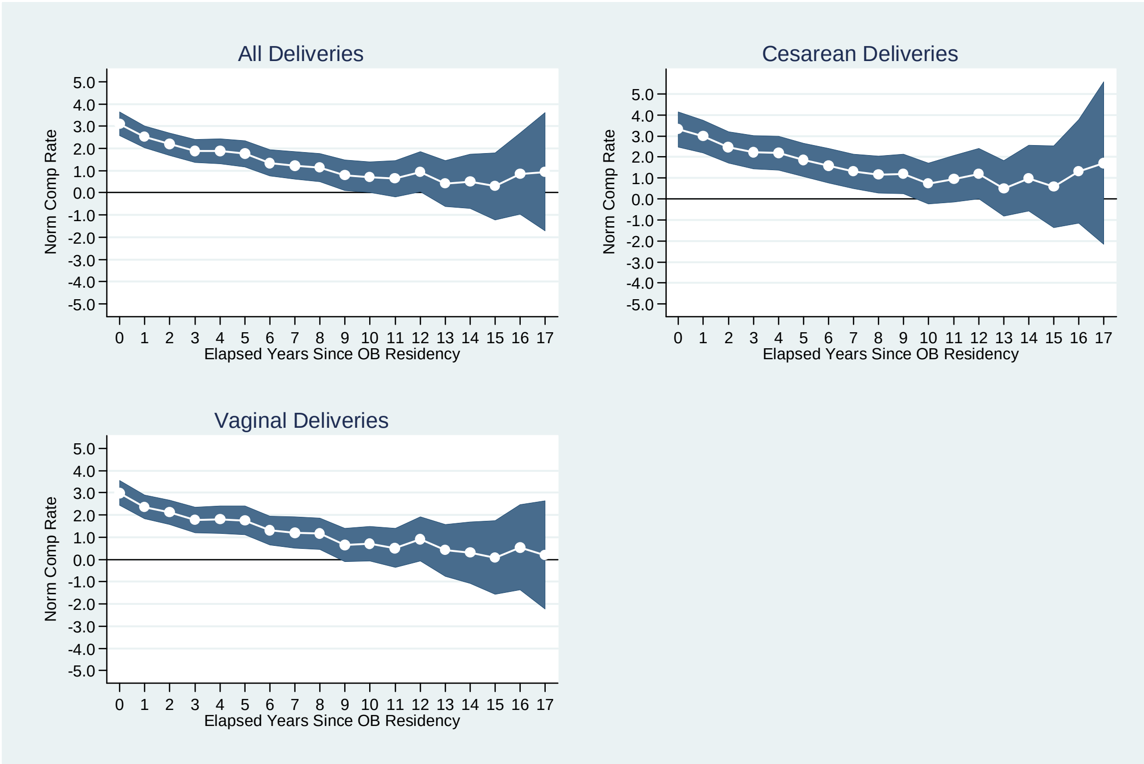
Wooldridge, J.M., 2006. Introductory Econometrics: A Modern Approach, 3rd Ed. Thomson-SouthWestern: Mason, OH.

Table 1. Sample descriptive statistics at the physician-year and physician levels

	Physician-Years		Physicians	
	N	%	N	%
Total	15,673	--	1,864	--
State				
Florida	4,864	31.0	604	32.4
New York	10,809	69.0	1,260	67.6
Sex				
Male	7,351	46.9	732	39.3
Female	8,322	53.1	1,132	60.7
Specialty				
Obstetrics-Gynecology	15,232	97.2	1,818	97.5
Maternal-Fetal Medicine	441	2.8	46	2.5
Medical School Location				
United States or Canada	39,177	71.6	1,535	82.4
Other	15,559	28.4	329	17.6
Practice Duration				
No Deliveries in AY 2009	3,228	20.6	532	28.5
≥ 1 Delivery in AY 2009	12,445	79.4	1,332	71.5
Time Since OB-Gyn Residency Completion				
0-4 Years	7,979	50.9	--	--
5-9 Years	4,874	31.1	--	--
10-14 Years	2,414	15.4	--	--
15-17 Years	406	2.6	--	--
Duration in Sample				
1-5 Years	--	--	646	34.7
6-10 Years	--	--	549	29.5
11-15 Years	--	--	480	25.7
16-18 Years	--	--	189	10.1

AY = Academic Year

Figure 1. Physician performance trajectories by delivery mode



Each graph shows annual mean physician-level normalized risk-adjusted complication rates and 95% confidence intervals by elapsed years of experience since OB residency completion, as estimated by weighted least squares regression.

Table 2. Effects of years since residency on normalized risk-adjusted maternal complication rates

	WLS, All Docs		WLS, Stayer Docs	
	(1)	(2)	(3)	(4)
Fixed effects	None	Doc	None	Doc
All Deliveries				
Years since residency	-0.18***	-0.13***	-0.15***	-0.13***
	[0.023]	[0.025]	[0.024]	[0.026]
Constant	2.65***	2.05***	2.38***	1.92***
	[0.15]	[0.13]	[0.17]	[0.15]
R^2	0.012	0.575	0.009	0.566
N Doc-years	15,673		12,445	
N Docs	1,864		1,332	
Cesarean Deliveries				
Years since residency	-0.18***	-0.068*	-0.16***	-0.066*
	[0.030]	[0.036]	[0.032]	[0.038]
Constant	2.92***	1.90***	2.71***	1.70***
	[0.21]	[0.21]	[0.24]	[0.24]
R^2	0.006	0.513	0.005	0.511
N Doc-years	15,445		12,373	
N Docs	1,863		1,332	
Vaginal Deliveries				
Years since residency	-0.18***	-0.16***	-0.15***	-0.16***
	[0.024]	[0.025]	[0.025]	[0.026]
Constant	2.56***	2.07***	2.27***	1.95***
	[0.16]	[0.13]	[0.18]	[0.15]
R^2	0.010	0.549	0.007	0.540
N Doc-years	15,548		12,396	
N Docs	1,864		1,332	

WLS weights based on Borjas (1987). Robust standard errors clustered by physician in brackets.

*** p<0.01, ** p<0.05, * p<0.1

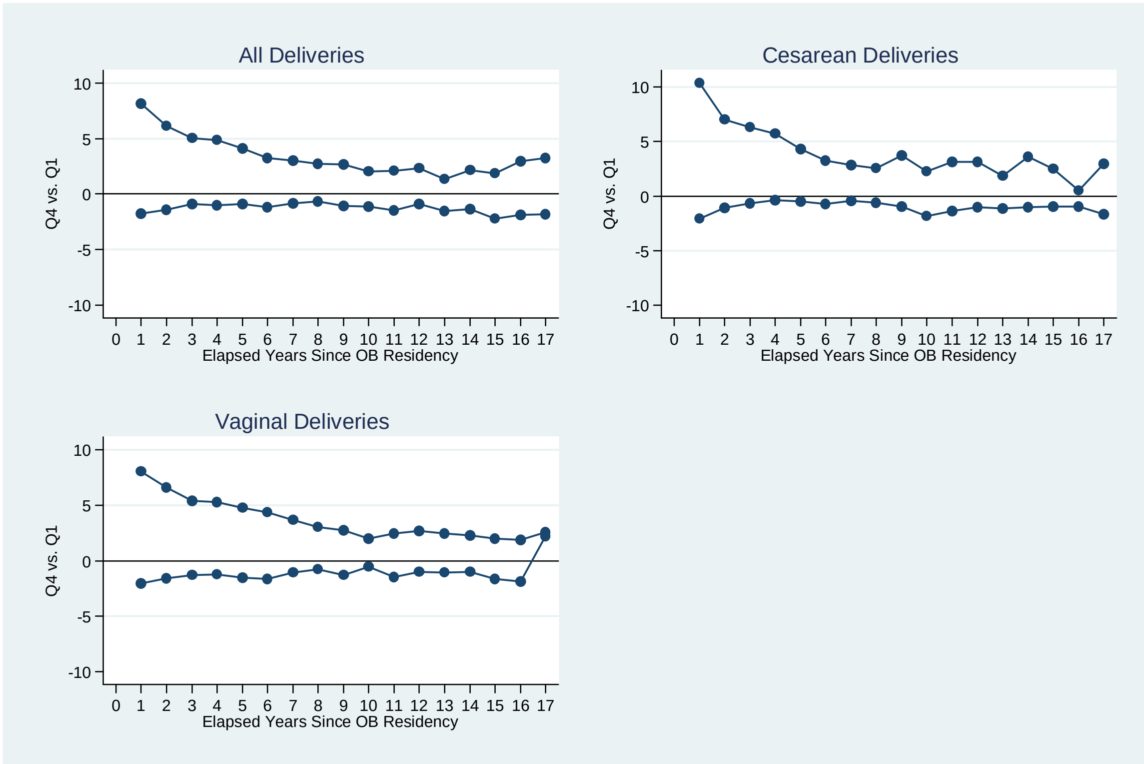
Table 3. Correlation of physicians' risk-adjusted complication rates over time

	Deliveries		
Lag	All	Cesarean	Vaginal
1	0.49***	0.42***	0.48***
	13,647	13,424	13,537
2	0.42***	0.35***	0.39***
	11,888	11,672	11,784
3	0.35***	0.27***	0.34***
	10,284	10,092	10,184
4	0.31***	0.25***	0.31***
	8,859	8,674	8,765
5	0.27***	0.21***	0.29***
	7,567	7,406	7,482
6	0.25***	0.19***	0.27***
	6,389	6,242	6,311
7	0.25***	0.17***	0.25***
	5,323	5,193	5,255
8	0.21***	0.16***	0.23***
	4,380	4,263	4,315
9	0.19***	0.14***	0.21***
	3,539	3,442	3,485
10	0.18***	0.15***	0.20***
	2,788	2,712	2,738
11	0.17***	0.11***	0.18***
	2,111	2,057	2,077
12	0.19***	0.13***	0.18***
	1,547	1,514	1,518
13	0.12***	0.14***	0.15***
	1,081	1,053	1,056
14	0.14***	0.16***	0.17***
	696	679	679
15	0.06	0.08	0.14***
	406	394	393
16	0.02	-0.03	0.13*
	202	197	193
17	0.03	0.14	-0.05
	64	62	62

Each cell displays product-moment correlation coefficient, significance and sample size, where the correlation is computed between current and lagged performance.

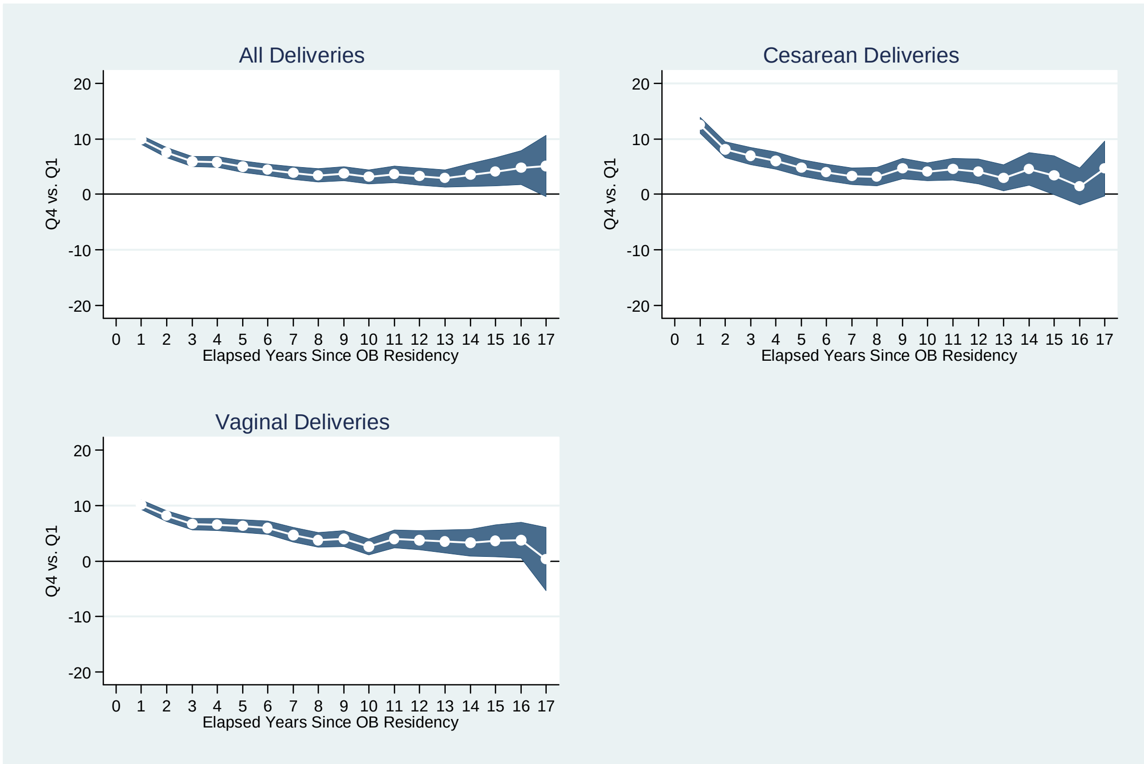
*** p<0.01, ** p<0.05, * p<0.1

Figure 2. Performance trajectories by delivery mode for physicians in the 1st and 4th quartiles of initial performance



Each graph shows annual mean physician-level normalized risk-adjusted complication rates by elapsed years of experience since OB residency completion for physicians in the 1st and 4th quartiles of performance in the first year after residency completion, as estimated by weighted least squares regression.

Figure 3. Differences in performance trajectories by delivery mode for physicians in the 1st and 4th quartiles of initial performance



Each graph shows arithmetic differences in annual mean physician-level normalized risk-adjusted complication rates and their 95% confidence intervals by elapsed years of experience since OB residency completion for physicians in the 1st and 4th quartiles of performance in the first year after residency completion, as estimated by weighted least squares regression with robust standard error clustered for multiple observations per physician.

Table 4. Sources of variation in physician performance

	All Years					
	All Docs	Annual Vol ≥ 30	Stayer Docs	Year 6	Year 11	Year 16
All Deliveries						
Total R^2	8.93	13.31	10.31	5.19	4.81	4.03
Skill	7.28	11.08	8.59	4.39	3.27	1.55
Contemporaneous volume	0.26	0.34	0.30	0.04	0.43	0.65
Lagged cum volume & experience	0.79	1.00	0.76	--	--	--
Cumulative volume	--	--	--	0.24	0.12	0.09
N Doc-Years	13,292	12,387	10,807	1,169	630	179
Cesarean Deliveries						
Total R^2	4.85	8.51	5.03	2.47	4.18	1.52
Skill	3.84	7.18	3.85	2.38	2.75	1.05
Contemporaneous volume	0.22	0.23	0.28	0.04	0.94	0.19
Lagged cum volume & experience	0.23	0.54	0.17	--	--	--
Cumulative volume	--	--	--	0.01	0.09	0.01
N Doc-Years	13,090	7,719	10,717	1,154	621	176
Vaginal Deliveries						
Total R^2	8.31	12.95	9.28	6.70	3.62	4.86
Skill	6.71	10.79	7.78	5.02	1.98	2.72
Contemporaneous volume	0.23	0.27	0.23	0.00	0.40	0.96
Lagged cum volume & experience	0.87	1.22	0.82	--	--	--
Cumulative volume	--	--	--	0.74	0.22	0.03
N Doc-Years	13,210	11,737	10,772	1,161	624	176

R^2 reported in units from 0 to 100 to enhance readability.
Cells indicate either total R^2 from model or unique contribution of specific factor(s) to R^2.
Each column × delivery mode block represents a distinct regression model.

Table 5. Sources of variation in physician performance

	All Years, All Docs		
	All Deliveries	Cesarean Deliveries	Vaginal Deliveries
Total R^2	21.66	20.36	19.63
Skill	1.95	0.5	2.17
Contemporaneous volume	0.13	0.02	0.14
Lagged cum volume & experience	0.55	0.13	0.69
Initial hospital fixed effects	12.73	15.51	11.32
N Doc-Years	13,292	13,090	13,210

R^2 reported in units from 0 to 100 to enhance readability.
Cells indicate either total R^2 from model or unique contribution of specific factor(s) to R^2.
Each column represents a distinct regression model.

Table 6. Effects of performance signal and prior on all delivery volume by years of physician experience

		OLS, All Docs			
		(1)	(2)	(3)	(4)
	Fixed effects	None	Market	Hospital	Doc
Net effect of Performance SIGNAL in...	2nd year	-1.87	-1.58	-2.46	0.072
		[1.73]	[3.21]	[2.22]	[3.15]
	3rd year	-5.99***	-5.96**	-6.17***	-0.95
		[2.21]	[2.09]	[2.05]	[2.05]
	4th year	-0.24	-0.60	-0.61	1.46
		[6.00]	[3.86]	[5.00]	[3.85]
	5th year	-3.50	-4.17	-5.30**	0.077
		[2.90]	[3.29]	[2.66]	[2.71]
	6th and 7th years	-1.06	-2.64	-4.02	-0.63
		[2.41]	[1.81]	[2.45]	[2.10]
	8th-12th years	-3.15	-4.77	-4.65*	-4.04***
		[2.37]	[2.82]	[2.44]	[1.54]
	13th-18th years	0.97	-1.66	-0.66	-0.22
		[4.04]	[3.28]	[4.01]	[2.89]
Net effect of Performance PRIOR in…	2nd year	--	--	--	--
	3rd year	1.31	1.37	0.84	-0.062
		[1.52]	[1.98]	[1.35]	[2.04]
	4th year	-2.66	-2.07	-3.22	-1.53
		[3.04]	[3.42]	[2.58]	[2.84]
	5th year	-2.86	-1.91	-1.95	-1.33
		[2.38]	[2.14]	[2.52]	[2.90]
	6th and 7th years	-3.03	-1.92	-0.88	0.87
		[2.67]	[1.91]	[3.05]	[3.00]
	8th-12th years	-8.56**	-8.74*	-7.06*	-2.43
		[3.63]	[4.38]	[4.10]	[3.81]
	13th-18th years	-17.2**	-18.1**	-15.9**	-7.38
		[7.51]	[5.95]	[7.40]	[6.51]

N=13,292 doc-years and 1,750 docs
*** p<0.01, ** p<0.05, * p<0.1
Coefficients reflect the impact on annual all delivery volume of a one standard deviation increase in the signal or prior. OLS models control for physician characteristics (sex, maternal fetal medicine specialty, international medical graduate status, practice at multiple hospitals, year of OB residency completion), hospital characteristics (annual average number of deliveries, OB residency sponsorship, OB residency affiliation), state, and academic year, except where collinear with fixed effects. Robust standard errors are adjusted for clustering at the physician level (models 1 and 4), market level (model 2), or hospital level (model 3).

Table 7. Effects of performance signal and prior on Cesarean delivery volume by years of physician experience

		OLS, All Docs			
		(1)	(2)	(3)	(4)
	Fixed effects	None	Market	Hospital	Doc
Net effect of Performance SIGNAL in...	2nd year	0.86*	0.29	0.60	1.37
		[0.47]	[0.48]	[0.60]	[0.88]
	3rd year	-0.15	-0.42	-0.54	0.015
		[0.54]	[0.50]	[0.56]	[0.54]
	4th year	-1.31**	-1.40***	-1.31**	-1.15**
		[0.62]	[0.37]	[0.63]	[0.48]
	5th year	-1.26*	-1.46**	-1.49**	-0.67
		[0.68]	[0.50]	[0.60]	[0.52]
	6th and 7th years	0.20	0.055	0.047	-0.40
		[0.65]	[0.56]	[0.64]	[0.47]
	8th-12th years	-2.10***	-2.32**	-2.09***	-2.31***
		[0.76]	[0.90]	[0.71]	[0.56]
	13th-18th years	-2.33	-2.72	-2.05	-0.79
		[1.81]	[2.24]	[1.61]	[1.10]
Net effect of Performance PRIOR in...	2nd year	--	--	--	--
	3rd year	0.36	0.052	0.28	0.47
		[0.37]	[0.32]	[0.40]	[0.58]
	4th year	0.67	0.23	0.36	0.59
		[0.48]	[0.53]	[0.51]	[0.72]
	5th year	0.13	-0.43	0.057	-0.065
		[0.72]	[0.63]	[0.71]	[0.83]
	6th and 7th years	-1.49*	-2.16***	-1.21	-1.59*
		[0.77]	[0.62]	[0.80]	[0.84]
	8th-12th years	-3.49***	-4.67***	-3.46***	-4.28***
		[1.34]	[1.32]	[1.18]	[1.28]
	13th-18th years	-4.16	-5.82*	-4.19	-5.92**
		[3.52]	[2.85]	[3.19]	[2.61]

N=13,101 doc-years and 1,744 docs
*** p<0.01, ** p<0.05, * p<0.1
Coefficients reflect the impact on annual Cesarean delivery volume of a one standard deviation increase in the signal or prior. OLS models control for physician characteristics (sex, maternal fetal medicine specialty, international medical graduate status, practice at multiple hospitals, year of OB residency completion), hospital characteristics (annual average number of deliveries, OB residency sponsorship, OB residency affiliation), state, and academic year, except where collinear with fixed effects. Robust standard errors are adjusted for clustering at the physician level (models 1 and 4), market level (model 2), or hospital level (model 3).

Table 8. Effects of performance signal and prior on vaginal delivery volume by years of physician experience

		OLS, All Docs			
		(1)	(2)	(3)	(4)
	Fixed effects	None	Market	Hospital	Doc
Net effect of Performance SIGNAL in...	2nd year	-3.56***	-2.61	-3.00*	-1.81
		[1.29]	[2.38]	[1.54]	[2.21]
	3rd year	-4.00***	-3.65*	-3.82***	-0.68
		[1.38]	[1.73]	[1.38]	[1.38]
	4th year	-0.60	-0.45	0.038	2.66
		[4.41]	[2.93]	[3.52]	[2.82]
	5th year	-1.04	-1.08	-1.21	1.29
		[2.10]	[2.98]	[2.17]	[2.12]
	6th and 7th years	-1.40	-2.42*	-2.84*	0.63
		[1.66]	[1.18]	[1.71]	[1.56]
	8th-12th years	-2.64*	-3.39**	-2.43	-1.31
		[1.48]	[1.28]	[1.56]	[1.02]
	13th-18th years	0.020	-0.55	-1.36	-0.68
		[2.50]	[1.66]	[2.59]	[1.99]
Net effect of Performance PRIOR in…	2nd year	--	--	--	--
	3rd year	-0.40	0.014	-0.31	-0.78
		[1.03]	[1.43]	[0.96]	[1.39]
	4th year	-2.71	-1.77	-2.51	-2.04
		[2.39]	[2.88]	[2.02]	[2.08]
	5th year	-3.68**	-2.33	-2.22	-1.08
		[1.74]	[1.93]	[1.84]	[2.05]
	6th and 7th years	-1.22	0.35	0.72	2.30
		[2.03]	[1.43]	[2.20]	[2.29]
	8th-12th years	-3.87	-3.21	-2.34	1.93
		[2.39]	[2.43]	[2.58]	[2.69]
	13th-18th years	-10.0**	-10.3***	-7.46*	-0.12
		[4.49]	[3.18]	[4.12]	[4.42]

N=13,216 doc-years and 1,749 docs
*** p<0.01, ** p<0.05, * p<0.1
Coefficients reflect the impact on annual vaginal delivery volume of a one standard deviation increase in the signal or prior. OLS models control for physician characteristics (sex, maternal fetal medicine specialty, international medical graduate status, practice at multiple hospitals, year of OB residency completion), hospital characteristics (annual average number of deliveries, OB residency sponsorship, OB residency affiliation), state, and academic year, except where collinear with fixed effects. Robust standard errors are adjusted for clustering at the physician level (models 1 and 4), market level (model 2), or hospital level (model 3).

Table 9. Effects of performance signal and prior on all delivery volume for patients with commercial insurance by years of physician experience

		OLS, All Docs			
		(1)	(2)	(3)	(4)
	Fixed effects	None	Market	Hospital	Doc
Net effect of Performance SIGNAL in...	2nd year	-1.30	-2.92***	-3.09***	-0.62
		[0.90]	[0.69]	[1.17]	[1.78]
	3rd year	-2.46*	-3.74**	-2.97**	-0.20
		[1.26]	[1.18]	[1.37]	[1.18]
	4th year	-3.79***	-4.16***	-1.79	-1.06
		[1.21]	[0.88]	[1.16]	[0.85]
	5th year	-2.49	-3.10*	-4.69***	-0.24
		[1.57]	[1.42]	[1.46]	[1.14]
	6th and 7th years	-4.78***	-5.50***	-4.48***	-1.60**
		[1.35]	[1.50]	[1.21]	[0.77]
	8th-12th years	-3.32**	-4.63***	-3.04**	-2.20**
		[1.48]	[0.88]	[1.25]	[0.92]
	13th-18th years	-0.22	-1.50	-0.10	-0.98
		[2.64]	[1.86]	[2.41]	[1.72]
Net effect of Performance PRIOR in…	2nd year	--	--	--	--
	3rd year	0.54	0.015	-0.66	0.011
		[0.87]	[0.81]	[0.86]	[1.15]
	4th year	0.36	-0.97	-1.82*	-0.36
		[1.03]	[0.74]	[1.03]	[1.47]
	5th year	-1.10	-2.55**	-0.98	-1.43
		[1.46]	[1.02]	[1.47]	[1.56]
	6th and 7th years	-1.01	-2.68	-2.49	-0.84
		[1.57]	[1.67]	[1.57]	[1.60]
	8th-12th years	-5.49**	-7.11***	-7.21***	-3.24*
		[2.16]	[2.16]	[1.91]	[1.85]
	13th-18th years	-5.34	-7.18	-8.47**	-3.49
		[3.99]	[4.10]	[3.57]	[3.34]

N=13,292 doc-years and 1,750 docs
*** p<0.01, ** p<0.05, * p<0.1
Coefficients reflect the impact on annual delivery volume among patients with non-Medicaid insurance of a one standard deviation increase in the all-delivery signal or prior. OLS models control for physician characteristics (sex, maternal fetal medicine specialty, international medical graduate status, practice at multiple hospitals, year of OB residency completion), hospital characteristics (annual average number of deliveries, OB residency sponsorship, OB residency affiliation), state, and academic year, except where collinear with fixed effects. Robust standard errors are adjusted for clustering at the physician level (models 1 and 4), market level (model 2), or hospital level (model 3).

Table 10. Effects of performance signal and prior on all delivery volume for patients with Medicaid or no insurance by years of physician experience

		OLS, All Docs			
		(1)	(2)	(3)	(4)
	Fixed effects	None	Market	Hospital	Doc
Net effect of Performance SIGNAL in...	2nd year	-0.57	1.33	0.62	0.69
		[1.60]	[3.12]	[1.78]	[2.49]
	3rd year	-3.53*	-2.22	-3.20**	-0.75
		[1.83]	[1.99]	[1.43]	[1.60]
	4th year	3.55	3.56	1.18	2.53
		[6.04]	[3.36]	[4.46]	[3.47]
	5th year	-1.01	-1.07	-0.60	0.31
		[2.45]	[2.52]	[2.01]	[2.17]
	6th and 7th years	3.72*	2.85	0.46	0.97
		[2.02]	[2.10]	[2.40]	[1.98]
	8th-12th years	0.18	-0.14	-1.61	-1.85
		[1.87]	[2.72]	[1.98]	[1.26]
	13th-18th years	1.19	-0.16	-0.56	0.75
		[3.14]	[2.58]	[2.77]	[2.74]
Net effect of Performance PRIOR in…	2nd year	--	--	--	--
	3rd year	0.77	1.36	1.50	-0.073
		[1.24]	[1.38]	[1.03]	[1.56]
	4th year	-3.01	-1.10	-1.40	-1.18
		[2.96]	[3.00]	[2.29]	[2.22]
	5th year	-1.76	0.64	-0.97	0.10
		[2.13]	[1.79]	[2.05]	[2.31]
	6th and 7th years	-2.02	0.75	1.61	1.72
		[2.39]	[1.82]	[2.66]	[2.49]
	8th-12th years	-3.07	-1.63	0.15	0.82
		[3.24]	[3.37]	[3.46]	[3.23]
	13th-18th years	-11.9**	-11.0**	-7.40	-3.89
		[5.93]	[4.91]	[5.24]	[5.57]

N=13,292 doc-years and 1,750 docs
*** p<0.01, ** p<0.05, * p<0.1
Coefficients reflect the impact on annual delivery volume among patients with Medicaid or no insurance of a one standard deviation increase in the all-delivery signal or prior. OLS models control for physician characteristics (sex, maternal fetal medicine specialty, international medical graduate status, practice at multiple hospitals, year of OB residency completion), hospital characteristics (annual average number of deliveries, OB residency sponsorship, OB residency affiliation), state, and academic year, except where collinear with fixed effects. Robust standard errors are adjusted for clustering at the physician level (models 1 and 4), market level (model 2), or hospital level (model 3).

Table 11. Effects of performance signal and prior on gynecological volume by years of physician experience

		OLS, All Docs			
		(1)	(2)	(3)	(4)
	Fixed effects	None	Market	Hospital	Doc
Net effect of Performance SIGNAL in...	2nd year	0.15	-0.64	-0.56*	0.23
		[0.25]	[0.45]	[0.29]	[0.40]
	3rd year	0.39	-0.28	0.20	0.48*
		[0.31]	[0.31]	[0.30]	[0.28]
	4th year	-0.0046	-0.36	0.059	0.31
		[0.35]	[0.37]	[0.32]	[0.25]
	5th year	-0.0058	-0.68	-0.28	-0.15
		[0.47]	[0.45]	[0.40]	[0.28]
	6th and 7th years	0.18	-0.40	0.051	0.30
		[0.34]	[0.33]	[0.26]	[0.21]
	8th-12th years	0.62*	-0.23	0.14	-0.07
		[0.34]	[0.23]	[0.28]	[0.17]
	13th-18th years	0.90	-0.45	0.057	0.55*
		[0.71]	[0.51]	[0.56]	[0.32]
Net effect of Performance PRIOR in…	2nd year	--	--	--	--
	3rd year	0.070	-0.23	-0.39*	-0.03
		[0.22]	[0.39]	[0.20]	[0.28]
	4th year	0.24	-0.39	-0.42	0.046
		[0.36]	[0.50]	[0.36]	[0.36]
	5th year	-0.088	-0.63	-0.56	0.050
		[0.41]	[0.56]	[0.34]	[0.37]
	6th and 7th years	-0.11	-0.96	-0.71*	-0.05
		[0.41]	[0.58]	[0.43]	[0.39]
	8th-12th years	-1.08*	-2.05**	-1.27**	-0.17
		[0.57]	[0.75]	[0.54]	[0.47]
	13th-18th years	-1.46*	-2.25**	-1.05	-0.10
		[0.87]	[0.99]	[0.69]	[0.74]

N=13,292 doc-years and 1,750 docs
*** p<0.01, ** p<0.05, * p<0.1
Coefficients reflect the impact on annual gynecological procedure volume (oophorectomies and hysterectomies) of a one standard deviation increase in the all-delivery signal or prior. OLS models control for physician characteristics (sex, maternal fetal medicine specialty, international medical graduate status, practice at multiple hospitals, year of OB residency completion), hospital characteristics (annual average number of deliveries, OB residency sponsorship, OB residency affiliation), state, and academic year, except where collinear with fixed effects. Robust standard errors are adjusted for clustering at the physician level (models 1 and 4), market level (model 2), or hospital level (model 3).

Table 12. Effects of performance signal and prior on hazard of exit by years of physician experience

	Signal	Prior
3rd year	1.10*	0.13
	[0.66]	[0.33]
4th year	-0.39	0.17
	[1.06]	[0.57]
5th year	-0.40	0.71
	[1.49]	[0.86]
6th and 7th years	0.57	0.38
	[0.65]	[0.55]
8th-12th years	0.72	0.24
	[0.63]	[0.59]
13th-18th years	0.43	-0.41
	[1.47]	[1.32]

N=11,982 doc-years and 1,670 docs
*** p<0.01, ** p<0.05, * p<0.1
Average marginal effects reflect the impact on the hazard of exit (shown in percentage point terms) of a one standard deviation increase in the all-delivery signal or prior. Complementary log-log models control for physician characteristics (sex, maternal fetal medicine specialty and international medical graduate status), state and academic year. Standard errors are robust and clustered at the physician level.