THE SUBSTANCE OF STYLE: A STUDY OF SMALL AREA VARIATIONS IN THE PRACTICE STYLES OF OB/GYN SPECIALISTS

David Dranove*

Subramaniam Ramanarayanan**

Hayagreeva Rao¶

Draft: Please do not cite without permission of authors

Abstract

We propose several new statistical methods for assessing the presence and sources of small area variations in physician practice. Applying these methods to the choice of caesarian versus vaginal delivery in Florida, we show that half of the observed variation in regional practice is a statistical artifact resulting from variation in individual physician practice styles. The remaining half is due largely to variation in practice styles across hospitals. We further show that hospital-level variation results from physician sorting, as "doctors of a feather" practice together, which in turn stems from variations in medical training. Patient selection of physicians also contributes to observed variations.

* Kellogg School of Management, Northwestern University

** UCLA Anderson School of Management

¶ Stanford University Graduate School of Business

I. Introduction

Since Wennberg and Gittelson (1973) demonstrated that a patient's likelihood of receiving a tonsillectomy varied from 7% to 70% across similar towns in Vermont, numerous studies have shown that the type of treatment a patient receives may depend on where the patient lives, and not just on what condition the patient has. (See Phelps and Mooney, 1993 for detailed reviews.) Much of this research considers variations across geographical locales (intra-state or inter-state) and finds that the rates of variation as measured by the coefficient of variation (CoV) range from less than .15 for well-understood procedures such as hip-fractures to CoV's higher than .50 for poorly understand procedures such as treatments for back-injury or diabetes (Phelps, 1999).

Many scholars have argued that such small area variations (SAVs) in practice styles[1] imply market failure and lead to welfare losses. For instance, Phelps and Parente (1990) estimated that the annual welfare loss in 1987 due to small area variations was $33 billion dollars and noted that this was an understatement if there is variation within a market and between markets. Health care economists have offered a plethora of solutions to remedy them ranging from standardizing medical curricula to altering incentives. The effectiveness of particular solutions would depend, naturally, on the reasons why SAVs exist, which is the subject of this analysis.

A number of traditional explanations for SAVs such as differences in patient preferences, income, underlying health status of physicians, physician density, access to medical care, availability of substitutes, or noise stemming from sampling have been disconfirmed in the literature (Phelps and Mooney, 1993; Wennberg and Gittelson, 1982; Bhikchandani, Chandra, Goldman, and Welch, 2001). Other writers have suggested that physician learning leads to convergence around community norms (Phelps and Mooney, 1999) but a recent study by Epstein and Nicholson (2005) shows that variation within a market is two to three times greater than variation across markets, and suggest that there is

[1] We use the term practice style to connote the propensity of a physician to treat a given patient in a particular way. For example, some physicians may have a propensity to deliver babies by caesarian section whereas others may favor vaginal delivery, all else equal.

little revision of prior beliefs by physicians. Burke, Fournier and Prasad (2004) note that "while there is a vast literature describing the phenomenon, the puzzle itself largely remains unresolved".

SAVs are also an anomaly for sociologists, but have received scant attention (Flood and Fennel, 1997). Cultural-frame institutionalists suggest that social systems are likely to exhibit uniformity of structures and styles when individuals are subjected to either coercive pressure from regulators, normative pressures from professional networks, or mimetic pressure from peers (DiMaggio and Powell, 1983; Scott, 2001). The managed care model underlying the organization of medical care in the U.S. reflects a tight integration of regulative, normative and mimetic influences (Scott, Ruef, Coronna and Mendel, 2000). So the persistence of SAVs is an anomaly, and hence, an opportunity to study whether social diversity is based on spatial boundaries in a profession where there is a cultural pressure towards homogeneity. Medical sociologists have studied disparities in health care caused by race and status (Mirowsky, Ross and Reynolds, 2000), but a pressing challenge for medical sociologists is to understand "what might account for the large amount of unwarranted variation" in clinical practices (Shortell, 2004:14).

These considerations provide the motivation for our study. We begin by documenting the regional CoV in caesarian section rates across counties in Florida. We then assess whether SAVs are artifacts of individual-level variations among physicians, or artifacts of hospital-level variations, and address this issue using a simple but novel method that can be easily generalized to other procedures/regions. We specifically ask "Is the measured CoV when physicians and their patients are grouped by predetermined geographic region larger than the CoV obtained when physicians and hospitals are randomly assigned to "pseudo" regions that are equal in number and makeup to the actual regions?" We find that the median CoV in the actual data is roughly comparable to the median CoV's obtained through the random assignment of hospitals to pseudo-counties. This suggests that when it comes to c-sections in Florida, there do not appear to be any meaningful *regional* practice style effects; all variations in practice style are limited to variations across *hospitals*.

We then ask why are physician styles correlated by hospital. We focus on three major explanations. First, we test for matching of physicians to hospitals on the basis of style.

Here again, we undertake a simple but novel approach that is easily generalized to other procedures/regions. We predict a physician's style on the basis of a vector of observable physician characteristics, including medical school and residency training identifiers. We correlate predicted styles of new physicians in year T, where T is the year in which they joined the hospital, with the actual styles in year T-1 of the hospitals that they joined, and find evidence of a substantial *matching effect*. (We think of this as a "doctors of a feather practice together" phenomenon.) We then test to discern if physician styles evolve over time – a *mimicry* or *learning* effect – and find them to be fairly stable. In particular, we find evidence in support of learning only in the short term, i.e. over the course of the first year a physician works at a hospital. This suggests that physician styles are imprinted early on in the careers of physicians and persist due to matching with the hospital rather than learning.

Finally, we find evidence that patients select physicians whose styles best match their own specific needs; i.e., a patient who is likely to require a caesarian will tend to select a physician whose style favors performing caesarians. This patient *selection* effect intensifies the measured SAVs that result from the physician matching described above.

II. Literature Review

The prevalence of SAVs in a wide range of medical specialties has occasioned a number of explanations. Early explanations hinged on variation in patient preferences. A study by Pritchard et al. (1998) refutes this argument. They asked patients at five medical centers their preferences for end-of-life care and found that the actual treatments were explained by the regional use of the form of care more than their preferences or clinical presentations. Another set of explanations implicate income and price variability across regions, but as Phelps (1999) notes, SAVs also exist in Canada where there is universal health care coverage. Yet another explanation is that SAVs are the outcome of random deviations from identical practices across communities, however, McPherson et al. (1981) reported that only 1-4 percent of the observed variation in Canada was due to noise.

Variations in illness patterns across regions are an intuitive explanation, but Fisher et al.'s (1994) study of a cohort of Medicare beneficiaries in the Boston area found wide variation across teaching hospitals, and little relationship between morbidity and actual hospitalization. Explanations hinging on access to medical care are also fragile because studies such as Bhikchandani et al. (2001) have shown that the CoV for myocardial infarction is lower than that for angina pectoris despite the fact both procedures require similar surgical and medical resources. Finally, variation in access to substitute procedures is another possible source of SAVs, but Phelps and Mooney (1993) show that the correlations among substitute procedures for ailments ranging from back-injury to strokes are positive rather than negative, thereby, implying that it is variation in physician perceptions rather than access to substitutes that matters.

Given the doubt cast on these traditional explanations, health care researchers have focused on the cultural foundations of spatial boundaries. Phelps and Mooney (1993) and Phelps (1999, 2003) postulate that physicians form beliefs about appropriate care during their medical and residency training, but learn from colleagues through Bayesian updating, and as a result, there is convergence around community norms. Since physicians with different training backgrounds will locate unevenly, there is likely to be inter-market variation in styles. Burke, Fournier and Prasad (2004) construct a formal model in which physician choices are shaped by a desire to conform to peers or spillovers of knowledge, and they show that small regional differences in patient mix give rise to different treatment patterns; so younger patients are less likely to receive bypass surgery when they live in an area where the average age of the population is higher.

There is some evidence for the fact that there is variability in resource use across medical schools in the U.S. Wennberg et al (2004) utilize claims data to document extensive variation across 77 Academic Medical Centers (AMCs) in the amount of care provided to Medicare FFS patients with three common conditions. They organize their analysis at the level of the hospital but acknowledge that practice styles of physicians, even those working at the same hospital, can differ and that this difference needs to be accounted for. If physicians develop their practice styles during medical school and residency training, variation across medical schools could act as a source of variation across physicians.

A recent study by Epstein and Nicholson (2005) of obstetrics and gynecology specialists in Florida shows that the variation in c-section rates across physicians within a market in Florida is two to three times greater than variation between markets. This suggests that physician-level variation may underlie regional variation. It also suggests that physicians do not substantially revise their beliefs due to local exchange of information, and as a result, physicians are unlikely to converge to a community standard, thus within-market variation is likely to persist. Epstein and Nicholson (2005) also report that residency training only explains four percent of the variation in risk-adjusted c-section rates across physicians, and note that although physicians learn from their peers, they do not significantly revise their beliefs about appropriate care.

Although Epstein and Nicholson's (2005) study is a valuable contribution, a number of challenges remain in understanding the social sources of SAVs. One set of issues concerns the level at which variation occurs; Epstein and Nicholson do not adequately consider a level of variation intermediate to the physician and the region, namely, the hospital. A corollary is whether SAVs at the regional level are a statistical artifact of variation at the hospital level.

A second set of issues pertains to whether the correlation among the styles of physicians in hospitals is simply an outcome of patient selection of physicians on the basis of unobservable patient characteristics (Burke, Fournier and Prasad, 2004), learning by physicians, or matching between physicians and hospitals. For instance, Epstein and Nicholson (2005) attempt to measure the extent to which physicians learn from their peers by regressing current practice style for each physician on the past practice styles of each physician's peers, and obtain a positive coefficient and infer that physicians are learning. Unfortunately, a positive coefficient is also consistent with selection and matching. We introduce methods that allow us to identify whether any or all of these phenomena are present.

Finally, we note that prior studies rely on hospital diagnostic coding to control for patient characteristics when measuring physician practice style. These studies neglect how variations in the coding practices and skills of hospital clerks may create hospital level correlations in "style" that belie the actual variations in physician practices. Thus, we rely on "transparent" patient characteristics that are unlikely to be subject to discretionary coding.

While the use of transparent coding (as opposed to all available diagnostic information) does not affect our main results, failure to take this step with other data could introduce bias.

III. Overview of our Approach

Because we are offering a number of new analytic techniques, it is useful to lay out our overall approach before delving into details. We begin by presenting a statistical model that lays the foundation for the subsequent work. This model demonstrates how unobservable physician and patient characteristics can contribute to measures of SAVs. Our subsequent empirical analyses are designed to avoid the resulting ambiguities in interpretation.

Our first empirical task is to determine if measured CoVs are artifacts of physician or hospital level variations. We begin by computing the actual CoV for c-sections across counties in Florida. Our choice of county as the regional unit is purely for analytic convenience. We next compute the empirical distribution of CoVs when physicians and hospitals are randomly assigned to "pseudo" regions that are equal in number and makeup to the actual regions. If the actual CoV exceeds the 95th percentile of the distribution of CoVs using pseudo-data, we conclude that there are genuine regional effects. Otherwise, we conclude that observed regional effects are artifacts of variation at a lower level of aggregation.

We then examine a series of possible explanations for CoVs. The most important of these is physician matching based on prior preferences. We could look at the correlation of physician styles within hospitals, but we would not be certain whether a positive correlation reflected matching based on prior preferences or other factors such as learning or the desire of physicians to conform to norms. To rule out the latter two explanations, we estimate a model predicting a physician's style on the basis of a vector of observable physician characteristics, including medical school and residency training identifiers. We then correlate predicted styles of new physicians in year T, where T is the year in which they joined the hospital, with the actual styles in year T-1 of the hospitals that they joined. We then examine whether the correlation increases over time to assess the extent of conforming/learning.

We conclude our empirical analysis by looking for evidence of patient sorting. We ask whether patients whose observable characteristics would indicate a preference for c-sections (e.g., they may have had a prior c-section) seek out physicians whose practice styles favor performing c-sections.[2]

We select Florida for our study of SAVs because the state makes available data that permits us to identify the physicians who perform surgical procedures. The Florida data is provided by the State Agency for Health Care Administration (AHCA). This data is comparable to patient level discharge data provided by the Health Care Utilization Project (HCUP) and the California Office of Statewide Health Planning and Development (OSHPD). Specifically, AHCA provides information about each discharge, including the hospital, diagnostic information, and demographic information (including the residence zip code.) In addition, AHCA provides the license number of the operating physician. We were able to obtain information about each physician's training and draw inferences about the physician's sex by linking the license to a separate online data base provided by the state of Florida. To protect confidentiality, we do not report any physician-specific information.

We study deliveries because there are many observations and there is a clear dichotomous decision – vaginal versus caesarian section – that may be used to identify practice "styles." A number of previous studies of practice variations have also examined the vaginal/caesarian dichotomy. We use data from 1994-2003 for various aspects of our study, but for convenience we document the extent of variations using only data from 2000-2001.

We consider each of the 67 counties in Florida to be a distinct market area. Previous studies of SAVs have used the county as the geographic unit. Using the county also allows us to focus on measuring the sources of variation, without dwelling on issues of market definition. As the main purpose is to suggest new analytic tools that can be applied to any

[2] In a clever study, Epstein and Nicholson (2005a) use information on practice variations in weekdays versus weekends as evidence of patient sorting.

procedure and market, we will not comment further on the appropriateness of using the county as the relevant geographic unit.

IV. A Statistical Model

The typical analysis of SAVs examines both raw and regression-adjusted procedure rates. We can represent this analysis using a standard binomial choice framework. We suppose that the decision to have a medical procedure is the result of a decision making process that incorporates both patient characteristics (e.g., the patient's clinical condition) and physician characteristics (e.g., the physician's training and beliefs). Let the benefits of a procedure *B* (relative to an alternative intervention) be represented by:

$$(1) \qquad B = B_0 + B_1X_p + B_2X_d + \varepsilon_p + \varepsilon_d$$

where X_p are observable (to the researcher) patient characteristics, X_d are observable physician characteristics, and ε_p and ε_d are unobservable patient and physician characteristics respectively. The parameters B_1 and B_2 measure the importance of observable physician and patient characteristics.

The patient undergoes the procedure if *B* exceed some cost *C* (that may include financial cost as well as inconvenience, etc.) Thus, the patient has the procedure if

$$(2) \qquad \varepsilon_p + \varepsilon_d > C - (B_0 + B_1X_p + B_2X_d)$$

This can be estimated using a logit or linear probability regression model:

$$(3) \qquad Y = f(X_p. X_d)$$

where the dependent variable Y = 0 if the patient has a vaginal delivery and Y = 1 if the patient has a caesarian.

In practice, analysts do not include physician characteristics in their estimate of equation (3). Instead, the typical regression model takes the form:

$$(4) \qquad Y = f(X_p)$$

The magnitude of SAVs is computed from estimates of equation (4) roughly as follows. Let e_i represent the regression error for patient i. Let R_{ej} represent the average value of e_i for all patients residing in a region j. (Alternatively, one can include region fixed effects in the regression and use the fixed effects coefficients to measure the average value for each region.) Let M equal the mean intervention rate in the population. Then the coefficient of variation (CoV) of regression-adjusted practice styles is given by the standard deviation of R_{ej} divided by M. (Alternatively, it is the standard deviation of the county fixed effects divided by M.) The CoV is the measure of SAV used by most researchers.

The CoVs that result from estimation of equation (4) necessarily incorporate variation in all physician characteristics as well as variation in unobservable patient characteristics. Thus, if $\varepsilon_p = 0$ (i.e., there are no unobservable patient characteristics that influence the value of a caesarian), then the measured SAVs will reflect only variations in physician characteristics. Now suppose $\varepsilon_p \neq 0$. Measured SAVs will reflect variations in physician characteristics *as well as variations in unobserved patient characteristics.* If ε_p is uncorrelated with X_d and ε_d, then the presence of unobserved patient characteristics will simply add noise to the measured practice variations and no harm is done by ignoring it. If ε_p is positively correlated with X_d and ε_d, then this correlation will intensify the measured extent of practice variations. Such a positive correlation would occur if patients with unobserved preferences for caesarians seek out physicians who themselves prefer to perform caesarians. This is certainly a plausible supposition.

To summarize, most studies of SAVs control for observable patient characteristics. Thus, the reported CoVs reflect intra-region variation in all physician characteristics and intra-region variation in correlated but unobservable patient characteristics.

V. Preliminary Analysis

In this section we document the extent of practice variations for c-sections in Florida during 2000 and 2001.[3] We begin by computing the county-level coefficient of variation (CoV) of "raw" practice styles. To do so, we simply find the mean and standard deviation of raw caesarian section rates in each county.

[3] Throughout this analysis, we restrict attention to physicians who perform at least 20 deliveries in the two years combined. See footnote 11 for further explanation.

Variations in raw caesarian section rates might be attributable to differences in patient characteristics across market areas. To account for this, we estimate equation (4) using a linear probability model. To compute the "fully adjusted" CoV, we include among the predictors both patient demographics (age, income, race, insurance status and a vector of county dummies) and numerous clinical indicators (including whether the mother had a previous caesarian and whether hypertension was listed as a secondary diagnosis). We also include the medical liability insurance rate for that county as a regressor, in order to capture the effect of malpractice pressures on caesarian section rates. [4] To compute the *county-level* CoV of fully adjusted style, we compute the CoV of the county fixed effects.[5]

The coding of some of the clinical indicators used to adjust practice style may be subjective. If so, the reporting of these indicators could vary by hospital for reasons that had nothing to do with the patient's health or physician's style. Referring back to our model, this would imply that X_p is measured with noise that is correlated among physicians within a given hospital. This would introduce correlation in the measured practice styles of physicians within a hospital and could bias upwards the measured CoV. Thus, we also compute a "partially adjusted" CoV in which we restrict the predictors to demographic variables and a handful of unambiguous clinical conditions (specifically, previous caesarian and multiple gestation) that we believe are likely to be consistently coded across hospitals.[6]

Table 1 reports the county-level CoVs of the unadjusted, partially adjusted, and fully-adjusted practice styles. We obtain a county level CoV of "raw" practice styles of 0.189, which is in line with previous estimates in the literature. The partially adjusted and fully adjusted CoVs are lower than the "raw" CoV, since we control for differences in patient characteristics across counties while computing these estimates. The fully adjusted CoVs are lower still, suggesting that controlling for all patient characteristics reduces observed variability, despite the potential for bias mentioned above. Even so, we will work with the

[4] We obtain data on malpractice insurance premiums for counties in Florida from the Medical Liability Monitor Annual Rate Survey. This survey divides Florida into three regions.

[5] We specifically compute each individual county's caesarian section rate at the mean of all predictors and compute the resulting CoV.

[6] We spoke with a "coding consultant" who works with hospitals to improve the accuracy of their coding and she confirmed that these codes are likely to be recorded consistently.

partially adjusted CoVs in future analyses, preferring noise to bias. None of our key findings are materially affected by this choice.

VI. Are SAVs Statistical Artifacts?

Physicians have their own individual practice styles.[7] It follows that there would be measurable (though possibly small) CoVs in caesarian rates across counties even if physicians were randomly allocated across the counties. By the same token, inter-hospital variation in practice styles will appear as county-level variations, even if hospitals are randomly located. It follows that an important step in assessing the magnitude of practice variations is to determine the level of aggregation at which variations occur.

The levels of aggregation we consider, ordered from lowest to highest, are as follows[8]:

· *Individual variation*: Measured SAVs at the hospital or county level are an artifact of differences among individual physicians.

· *Hospital*: Physicians with similar styles practice at the same hospital. Measured SAVs at the county level are an artifact of differences among hospitals.

· *County*: Measured SAVs at the county level cannot be explained by differences at the hospital level. Styles across hospitals within a county are positively correlated.

County-level SAVs could be statistical artifacts of lower level variation whenever: (1) there are large differences in physician/hospital practice styles and (2) few physicians/hospitals account for a large percentage of the total procedures in some regions. If (1) and (2) hold, the large variations at the physician/hospital level will show up as county-level variation, even though there are not county-level effects. In general, there is no way to be certain *a priori* whether observed SAVs at the county level are artifacts of lower level variation. One must examine the data.

[7] Phelps et al. (1994) document enormous variation in practice "styles" from one physician to the next.
[8] We would like to also consider aggregation at the level of the group practice, but appropriate data are not available.

We offer a novel method for identifying the proper level of aggregation. To illustrate our method, consider a hypothetical situation in which there are substantial differences in practice styles across physicians and only a handful of physicians in each region. Specifically, suppose that half of all physicians perform an invasive procedure on 40 percent of their patients, all else equal. The other half performs the procedure on only 20 percent of their patients. If each region has four physicians, then we will observe the following distribution of procedure rates:

Rate of Intervention	**Distribution of Regions**
40%	6.25%
35%	25%
30%	37.5%
25%	25%
20%	6.25%

The coefficient of variation (CoV) for this data is 0.17. This is in line with the estimated CoV for caesarians in Florida as well as many other published estimates, but in this example it is entirely due to random locations of individual physicians; there are no regional effects.

Here is another way to think about it. The "region" is a potentially arbitrary way of grouping physicians and their patients. If region really does not matter, then grouping into regions should be no different, in a statistical sense, than any other arbitrary grouping. To determine if region really does matter, we ask the following question:

> Is the CoV when physicians and their patients are grouped by predetermined geographic region larger than the CoV when physicians are randomly assigned to "pseudo" regions that are equal in number and makeup to the actual regions?[9]

[9] Note that in order to perform this exercise, we must "attach" patients to the physicians who treat them. Thus, the CoV in pseudo regions reflects random variations in both physician characteristics and unmeasured patient characteristics.

Random Assignment of Physicians to Pseudo-Counties

We treat each county as a region and reference the county level CoVs as reported in Table 1. Using the same data, we then compute the CoV for randomly constructed "pseudo counties" – random collections of physicians and their patients that resemble the actual counties in terms of number of physicians and patients. In this exercise, we use data aggregated over 2000 and 2001, and restrict our attention to physicians who perform at least 20 procedures in the two years combined.[10] We construct pseudo counties as follows:

1) We assign physicians to volume "tiers" based on the number of procedures performed by the physician in a county[11]: 400+ procedures; 200-400 procedures; 100-200 procedures; 20-99 procedures.

2) We identify where each physician practices. Let N_{tc} represent the number of physicians in tier t who perform deliveries in county c.

3) We select the first county and for each tier t in that county, we randomly select N_{t1} of the physicians assigned to tier t. We assign these physicians to the first "pseudo county."

4) We repeat the exercise for each remaining county, sampling without replacement. Once we have done this for all 67 counties, every physician is assigned to a pseudo county.

In this way, we create 67 pseudo counties that correspond to the original 67 Florida counties in the sense that they have the same number of physicians in each volume tier. We then compute the CoV for the pseudo counties. We repeat this experiment 1000 times to obtain the distribution of CoVs that would be obtained if all variation was due to random locations of physicians. We compare the actual CoV to this distribution to assess whether the actual value is larger than what we would expect to observe due to random chance.

Table 2 reports the CoVs for the actual counties and summary statistics on the distribution of CoVs in the pseudo counties, based on 1000 repetitions of the method

---

[10] We do this for analytic convenience. These physicians account for over 98 percent of all deliveries. The few remaining physicians would not have a dramatic impact on the measured CoV.

[11] A physician operating in two counties is treated as two physicians. Around 4 percent of physicians in our data operate in two counties.

described above. Table 3 reports the same data, restricted to rural counties.[12] There are two important takeaways from Tables 2 and 3. First, the CoV in the actual data exceeds the 99th percentile CoV in the pseudo data. We conclude that practice variation is not entirely a statistical artifact of differences across physicians. That said, the median CoV in the pseudo data is roughly 60 percent of the CoV in the actual data. This suggests that a substantial portion of observed SAVs in the actual data is due to random locations of physicians, rather than to any genuine regional effects.

Random Assignment of Hospitals to Pseudo-Counties

Now consider an intermediate level of aggregation – the hospital. We will first document practice variations at the level of the hospital. We then ask whether the CoV when hospitals and their patients are grouped by county is larger than the CoV when hospitals are randomly assigned to pseudo counties that are equal in number and makeup to the actual counties.

To do this, we repeat the preceding exercise at the level of the hospital rather than the physician:

1) We classify hospitals into four volume tiers: <1500 procedures; 1500-3000 procedures, 3000-6000 procedures and >6000 procedures.

2) We identify hospitals in each tier in each county. Let H_{tc} represent the number of hospitals in tier t in county c.

3) We select the first county and for each tier t in that county, we randomly select H_{t1} of the hospitals from tier t. We assign these hospitals to the first pseudo county.

4) We repeat the exercise for each remaining county, sampling without replacement. Once we have done this for all 67 counties, every hospital is assigned to a pseudo county.

Tables 4 and 5 report our findings for all counties and rural counties respectively. Tables 4 and 5 show that the actual CoVs are only slightly larger than the median CoVs in the pseudo data; the actual CoVs are smaller than the 95th percentile CoVs. There does not

[12] Counties having 3 hospitals or fewer were designated as rural counties.

appear to be any meaningful region practice style effects; all variations in practice style are limited to variations across hospitals.

To summarize, there are measurable practice variations in the raw and adjusted county level data. These variations are partially explained by differences in practice styles at the level of the physician. The remainder of these variations is explained by differences in styles at the level of the hospital. There is no meaningful additional variation at the highest level of aggregation, the county.

VII. Hospital-level Practice Styles

The preceding analysis is consistent with the idea that physicians practicing at a given hospital have positively correlated practice styles. This can occur for a variety of reasons:

(1) Physician matching: Physicians and hospitals choose one another based on common practice styles.

(2) Physician learning: Physicians who are new to a hospital learn or adopt the practice styles of their colleagues.

(3) Patient selection: There is a positive correlation in the unobservable characteristics of patients who choose that hospital.

In this section and the next, we offer direct and indirect evidence of each phenomenon. We begin by deriving physician styles by estimating equation (4). Recall that the resulting styles could reflect physician characteristics or unobservable characteristics of their patients. We then conduct a number of tests to confirm or rule out each of the reasons for hospital-level correlations in style. While we are able to affirm or dismiss each reason, we can not completely parse them out. That is, we cannot definitely state "X percent of hospital style is due to matching, Y percent to learning, and 100-X-Y percent to selection."

The estimation of equation (4) is neither novel nor central to our analysis so we will merely summarize our findings.[13] A patient with a previous caesarian has a much higher

[13] The full set of regression results are available from the authors upon request.

probability of delivering by caesarian in her current delivery. Older women are slightly more prone to having caesarians, as are black and Hispanic women (in comparison to white women). Patients living in higher income zip codes have a slightly smaller probability of having a caesarian. Increases in liability premiums are associated with slightly higher caesarian rates. The large sample size assured statistical significance of all the aforementioned predictors.

Consistent with our earlier analysis that revealed hospital-level practice styles, we find that the correlation between each physician's own adjusted style and the styles of all other physicians at the same hospital is .418, which is significant at p<.001.[14] (We computed this using transparent covariates so as to avoid inducing a spurious correlation attributable to common medical record keeping.) This confirms that physicians within a hospital have similar styles.

Evidence on Physician Matching

The matching hypothesis holds that physicians match with hospitals that share the same style. To search for evidence of matching, we examine the practice styles of physicians who are new to a hospital and the styles of that hospital in the year they join. We say that a physician is new to a hospital in a given year if the physician performed at least 25 procedures in that hospital in that year and had never performed more than 5 procedures in that hospital in any prior year for which we have data. We identified 767 new physicians using these criteria. Of these, 152 appear to be "brand new" physicians – new to a hospital and less than two years removed from completion of residency training.

We find that the correlation in practice style between new physicians and the existing physicians on the staff of their hospital is .356, while the correlation in practice style between

[14] When computing correlations between physicians and their hospitals, we exclude from the hospital those patients treated by the physician in question. In addition, we include only those physicians who treat at least 25 patients in a given hospital/year, so as to exclude physicians whose practice styles are estimated with substantial noise.

"brand new" physicians and existing physicians is .376. Both correlations are highly significant and consistent with matching.[15]

These correlations are also consistent with two alternative hypotheses: (a) physicians immediately adopt the styles of their colleagues upon joining a new hospital (what one might call *mimickry* or immediate learning), and (b) there are correlated but unobservable characteristics of the hospital's patients. To find additional evidence of matching, we develop a model that predicts each physician's practice style based on "predisposing characteristics" – demographic and medical training variables.[16] If we assume that physician selects a medical school and residency program prior to determining which hospital they wish to practice at, then a positive correlation between predicted physician styles and actual hospital styles would be definitive evidence of matching. We will call this "training-based" matching for simplicity.

The dependent variable for this model is the physician's style computed using transparent covariates as described above. The unit of observation in our regressions is a physician-hospital-year triad. For this analysis, we restrict our attention to physicians who perform more than 25 procedures at a hospital in any given year. This results in a sample containing 9180 observations with information on 1041 physicians across 10 years.[17]

Our predictors include a vector medical school and residency training identifiers (for those schools and programs that trained at least 5 physicians), a foreign medical school indicator, sex[18] and year of completion of medical school[19], and an interaction between the decade of graduation and the residency program indicator[20]. While constructing the

[15] An alternative explanation for these findings is that new physicians immediately mimic the styles of their senior colleagues. We cannot fully rule out this alternative; however, the evidence on training that we present below further supports the matching hypothesis.
[16] Phelps (2003) and others have observed that training can also have an important impact through a variety of mechanisms.
[17] These are the observations for which all variables have non-missing values. We were unable to obtain information on training programs for a few physicians, and they were excluded from this analysis
[18] While we lacked data on the physician's sex, we were able to make informed guesses based on the physician's first name. We had categories for male, female, and indeterminate.
[19] In alternate specifications, we tried including the number of years since graduation as a predictor, but dropped that variable in favor of year of completion for reasons of fit.
[20] This interaction was computed for those residency programs that train at least 20 physicians.

interaction variables, we ensured that each "cell" was sufficiently populated in order to avoid perfect prediction.

The model also incorporates detailed information on each training program (residency and medical school). We include indicator variables for the location (i.e. state) of the training programs in order to capture regional effects on practice style[21]. We include the ranking of these training programs as a linear predictor.[22] Finally, we also include a dummy variable that indicates whether or not a training program is affiliated with a university. The model also includes year of training indicators (aggregated into ten year periods) to capture time-specific factors that may affect practice style.

The results of this regression appear in Table 6. The medical school and residency indicators are highly significant; where you train affects how you practice. Male physicians tend to perform more caesarians. Physicians who underwent training at medical schools in foreign countries do not have significantly different practice styles in comparison to their American counterparts. The other coefficients indicate that the characteristics of the residency programs seem to be strong predictors of practice styles, especially in comparison to characteristics of the medical school. Finally, unranked training programs do not seem to have a different practice style in comparison to ranked programs, on average.

We recovered from this regression each physician's predicted style. We correlate predicted styles of new (to the hospital) physicians in year T, where T is the year in which they joined the hospital, with the actual styles in year T-1 of the hospitals that they joined. This correlation is .177, which is significant at $p<.001$. The correlation for brand new physicians is even stronger: .211. It is not surprising that these correlations are smaller than those reported earlier, as they are derived indirectly from a model predicting styles. Even so, the significant correlations provide definitive evidence of training-based matching of physicians to hospitals.

[21] We lumped together states that had less than 10 physicians.

[22] We used data from residentphysician.com to determine ranking of residency programs – based on the total amount of grants and awards received. We obtained rankings of medical schools from US News' "America's Best Graduate Schools 2006" survey. Programs that do not appear in the rankings are assigned the median rank but are also assigned a dummy variable. This allows complete flexibility for the average impact of unranked schools.

Physician Learning/Mimicry

Hospital-level styles may also result from learning, as new physicians gradually adopt the styles of their peers. Epstein and Nicholson (2005) attempt to measure the extent of learning by regressing current practice style of each physician on the past practice styles of peers. They obtain a positive coefficient. While this is consistent with learning, it is also consistent with matching. In this section, we propose a simple alternative way to detect learning.

We again focus on physicians who are new to their hospital and begin by examining immediate learning, or mimicry. Recall that the correlation in the practice styles of physicians in their first year at a new hospital and the practice styles of their colleagues was .356 (and .376 for brand new physicians). If we restrict attention to just the first three months after arrival, as opposed to the first year, this correlation is only.255 (and .203 in the case of "brand new" physicians). This suggests that physicians do learn from (i.e., mimic) their colleagues over the course of their first year at a hospital. Learning/mimicry might occur even faster (i.e., in the first three months) but we cannot be certain.

We now turn to the question of whether new physicians continue to "learn" from their peers over time. We correlate the practice style of new physicians for the first four years after joining a hospital, restricting attention to those years in which the physician performs at least 25 procedures. Tables 7 and 8 report the correlations. The correlation *remains more or less constant over time*. Taken together with the previous results on first-quarter correlations, we conclude that there is evidence in support of mimickry/learning in the short term, but that practice styles tend to be fairly stable over longer periods of time.

Patient Selection

As our theoretical model suggests, practice variations can result from correlations between physician characteristics and unobservable patient characteristics. We might expect a patient who has personal preferences for a caesarian section to select a physician who is predisposed to perform caesarians, thereby increasing the observed CoV. [23] We cannot, of

[23] Alternatively, doctors with particular styles may seek patients with matching needs.

course, detect whether unobservable patient characteristics are correlated with physician characteristics. However, it stands to reason that if there are correlations between observable patient characteristics and physician characteristics, then there will be correlations with the unobservable patient characteristics and physician characteristics as well. Thus, we can provide indirect evidence that selection on unobservables occurs, even if we cannot definitely measure its magnitude.

To test for this, we take our initial estimates of equation (4) predicting whether a given patient would have a caesarian. We use the results to predict the probability of a caesarian for each patient. This is a measure of the patient's observable preference for a caesarian. We find that the correlation between observable patient preferences and the practice styles of the physicians they select is .068, which is significant at $p<.001$. Note that the set of available patient observables is limited. Patients surely know much more than we do about their own preferences, so the extent of patient selection may be much higher than the correlation of .068 would suggest. In any event, we infer from the correlation between patient observables and physician styles that there is likely to also be a correlation between patient unobservables and physician styles. If so, this will increase the measured extent of SAVs.

VIII. Discussion

Our study was motivated by calls to understand the large amount of variation in clinical practices (Burke, Fournier, and Prasad, 2004; Shortell, 2004:14). Our main goal is to introduce several new techniques for measuring the extent, level, and sources of practice variations. We illustrate these techniques by examining obstetrics services in Florida. We do not assert that any of our findings will generalize to other services or regions. However, all of our methods can be generalized.

When it comes to c-sections in Florida, there do not appear to be any meaningful regional practice style effects; all variations in practice style are limited to variations across hospitals. This finding suggests that spatial boundaries by themselves may not matter much unless they hinge on social boundaries – in the case of Florida, organizational boundaries among hospitals were the progenitors of clinical variations.

A second implication of our study is that organizational variations among hospitals result from physician sorting. We also find that physician styles do not evolve over time. This suggests that physician styles are imprinted early on their careers and persist due to matching with the hospital rather than learning. Our analyses also revealed that patients select physicians whose styles best match their own specific needs; i.e., a patient who is likely to require a caesarian will tend to select a physician whose style favors performing caesarians. This patient selection effect intensifies the measured SAVs that result from the physician matching described above.

Taken together, these results inform theory and policy. One unmistakable implication for theories of practice styles is the role of imprinting. Phelps (2003) and a number of other scholars have attested to the significance of training as an antecedent of practice styles. Our study suggests that training has a "lock-in effect" that is redolent of imprinting. The ethologist Lorenz (1935) defined imprinting as a stamping process and argued that it was different from learning in two respects. First, susceptibility to imprinting is confined to a very limited period, usually, early in an organism's life. Second, once developed, imprinting is irreversible and so early experiences have a lock-in effect on organisms. Our analyses suggest that early physician training at medical school and residency – the formative period of a physician's career leads to irreversibility in styles. In turn, the salience of imprinting suggests that models of physician styles ought not to overweight the learning from peers at the expense of early career influences.[24]

Although our study is situated in the healthcare domain, it also speaks to the literature on organizational sociology. In recent years, a number of neo-institutional sociologists have suggested that organizational homogeneity is the outcome of coercive pressures from regulators, the mimetic effects of peers, and normative influences of

[24] Imprinting thus means socialization at the inception of a career Physicians get socialized in the hospitals and practices they join but as our analysis that does not matter much. It is socialization in graduate school and residency programs that is decisive.

professions (DiMaggio and Powell, 1983; Scott, 2001). Clinical decision making entails regulators, peers and professional influences and is therefore, an interesting setting to understand the scope and magnitude of homogeneity. Out study shows that professional training leads to heterogeneity across hospitals than homogeneity and indicates that the effects of peer learning are weak and that heterogeneity exists despite the presence of regulators. So neo-institutional research in sociology needs to recognize that professional influences can spawn variety rather than homogeneity.

Finally, our results have important implications for health care policy. The study of practice variations is motivated by the view that they may be socially harmful. Once we examine the sources of variation, this normative conclusion is not so clear. A root cause of practice variations is physician training. It is not obvious whether it is socially desirable for all physicians to receive identical training. The implications for experimentation and innovation would have to be explored. Variation is intensified by physician matching. We can easily imagine that this matching is socially desirable, as it can facilitate peer evaluation and thereby promote quality within the organization. Patient sorting also intensifies variation. The fact that variations in obstetrics practice in Florida are a hospital-level phenomenon suggests that patients have choices of local provider practice styles; our evidence of patient sorting suggests that some patients avail themselves of their choices. This can hardly be a bad thing.

References

Bikhchandani, S., Chandra, A., Goldman, D. and Welch, I., 2001, “The Economics of Iatroepidemics and Quakeries: Physician Learning, Information Cascades, and Geographic Variation in Medical Practice”, *Mimeo*

DiMaggio, Paul J. and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields" *American Sociological Review* 48: 147-160.

Epstein, Andrew, Jonathan Ketcham, and Sean Nicholson, 2005, "The Welfare Effects of Physician Specialization: Do We Want Physicians to Practice Alike?", working paper.

Fisher, E. S., J. E. Wennberg, T. A. Stukel, J. S. Skinner, and S. M. Sharp. 1994. "Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven." New England Journal of Medicine 331 (15): 989-95.

Flood AB and Fennell ML. "Through the Lenses of Organizational Sociology: The Role of Organizational Theory and Research in Conceptualizing and Examining Our Health Care System." *Journal of Health and Social Behavior*. 36:154–169, 1995

Fournier, G., Prasad, K. and Burke, M., 2004, “Physician Social Networks and Treatment Variations in Coronary Inpatient Care”, *Mimeo,* Florida State University

K. McPherson et al., “Regional Variation in the Use of Common Surgical Procedures: Within and Between England and Wales, Canada and the USA,” *Social Science and Medicine* Vol 15A, pp 273-288.

Mirkowsky, J., Ross, C. and Reynolds, J., 2000, “Links Between Health Status and Social Status” in Chloe Bird, Peter Conrad and Allan Fremont eds., *Handbook of Medical Sociology, Fifth Edition*, Pp. 47-67.

Nicholson, S. and Epstein, A., "The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections", 2005, NBER Working Paper.

Phelps, C. et al., "Doctors have Styles—And They Matter!" University of Rochester Working Paper.

Phelps, C., 2000, "Information Diffusion and Best Practice Adoption" in *Handbook of Health Economics*, edited by Culyer, A. and J. Newhouse, Amsterdam: Elsevier

Phelps, C. 2003. "What's Enough? What's Too Much"? Annals of Internal Medicine, 18: 4: 348-349.

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

Phelps, C. and Parente, S., 1990, "Priority Setting for Medical Technology and Medical Practice Assessment", *Medical Care*, 28(8), 703-23

Pritchard, RS, Fisher ES, Teno JM et al, 1998, "Influence of patient preferences and local health system characteristics on the place of death", *Journal of American Geriatric Society*, 46: 1242-1250

Scott, R.W., Ruef, M., Mendel, P. and Caronna, C. (2000), "Institutional Change and Health Care Organizations: From Professional Dominance to Managed Care", University of Chicago Press, Chicago.

Scott. W. R. 2001. Institutions and Organizations, 2nd edition. San Francisco, CA : Sage.

Shortell, S. 2004. "Increasing Value: A Research Agenda for Addressing the Managerial and Healthcare Challenges Facing Healthcare Delivery in the United States", Medical Care Research and Review, 61: 3: 12S-30S.

Wennberg JE, Fisher ES, Stukel TA, Sharp SM, 2004. "Use of Medicare Claims Data to Monitor Provider Specific Performance among Patients with Severe Chronic Illness", *Health Affairs,* Web exclusive, pp.5-18

Wennberg JE, Gittelson A., 1973, "Small area variation in health care delivery", *Science*, 182, 1102-08

Wennberg JE, Gittelson A., 1982, "Variation in medical care among small areas", *Scientific American*, 246:4: 120-134.

Table 1: CoVs for Caesarian Sections in Florida

Risk –Adjusters Used	CoV
None	.189
Partial	.147
Full	.127

Table 2: CoVs in Actual and Pseudo Counties – Physician Level Aggregation

	Actual counties	Pseudo counties 50th percentile	Pseudo counties 95th percentile	Pseudo counties 99th percentile
Raw Data	.189	.108	.146	.171
Transparent covariates	.147	.090	.111	.127
Full covariates	.127	.072	.097	.103

Table 3: CoVs in Actual and Pseudo Rural Counties – Physician Level Aggregation

	Actual counties	Pseudo counties 50th percentile	Pseudo counties 95th percentile	Pseudo counties 99th percentile
Raw Data	.220	.127	.169	.191
Transparent covariates	.169	.103	.125	.142
Full covariates	.143	.080	.102	.117

Table 4: CoVs in Actual and Pseudo Counties – Hospital Level Aggregation

	Actual counties	Pseudo counties 50th percentile	Pseudo counties 95th percentile	Pseudo counties 99th percentile
Raw Data	.189	.166	.198	.209
Transparent covariates	.147	.125	.152	.165
Full covariates	.127	.118	.135	.144

Table 5: CoVs in Actual and Pseudo Rural Counties – Hospital Level Aggregation

	Actual counties	Pseudo counties 50th percentile	Pseudo counties 95th percentile	Pseudo counties 99th percentile
Raw Data	.220	.197	.226	.241
Transparent covariates	.169	.142	.176	.194
Full covariates	.143	.129	.148	.166

Table 6: Determinants of Physician practice style – Regression results

Predictor	**Coefficient**
Male	.0148*** *(.002)*
Foreign Medical School	-.0175 *(.044)*
Rank of Residency program	-.0012*** *(.00018)*
Rank of Medical School	.0005 *(.0004)*
Residency hospital affiliated with university	-.0112** *(.0042)*
Is the residency program ranked – indicator	-.0050 *(.0048)*
Is the medical school ranked - indicator	.0029 *(.0076)*
Residency program indicators	Yes
Medical school indicators	Yes
Residency program – state indicators	Yes
Medical school – state indicators	Yes
Adjusted R^2 statistic	***.1579***
Number of Observations	***9180***

Notes: *Regression also includes year indicators and indicators for year of graduation as predictors. Sample restricted to all physicians who perform at least 25 procedures at a hospital in a year between 1994 and 2003. Standard errors in parentheses.*

**** - p < .001, ** - p<.01 and * - p<.05*

Table 7:Correlation between Physician Style and style of other physicians in Hospital

Note: t=0 denotes year of entry into hospital

	t=0	t=1	t=2	t=3	t=4
Correlation Coefficient	.3564	.3601	.3307	.3506	.3979

Table 8: Correlation between Physician Style and style of other physicians in Hospital – Brand New Physicians

Note: t=0 denotes year of entry into hospital

	t=0	t=1	t=2	t=3	t=4
Correlation Coefficient	.3763	.5501	.3541	.4254	.3736