MANAGEMENT SCIENCE
Vol. 52, No. 4, April 2006, pp. 473–488
ISSN 0025-1909 | EISSN 1526-5501 | 06 | 5204 | 0473

DOI 10.1287/mnsc.1050.0464


# The Firm Specificity of Individual Performance: Evidence from Cardiac Surgery

Robert S. Huckman
Harvard Business School, Boston, Massachusetts 02163, and National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, Massachusetts 20138, rhuckman@hbs.edu

Gary P. Pisano
Harvard Business School, Boston, Massachusetts 02163, gpisano@hbs.edu

In many settings, firms rely on independent contractors, or freelancers, for the provision of certain services. The benefits of such relationships for both firms and workers are often understood in terms of increased flexibility. Less understood is the impact of freelancing on individual performance. While it is often presumed that the performance of freelancers is largely portable across organizations, it is also possible that a given worker's performance may vary across organizations if he or she develops firm-specific skills and knowledge over time. We examine this issue empirically by considering the performance of cardiac surgeons, many of whom perform operations at multiple hospitals within narrow periods of time. Using patient mortality as an outcome measure, we find that the quality of a surgeon's performance at a given hospital improves significantly with increases in his or her recent procedure volume at *that* hospital but does not significantly improve with increases in his or her volume at *other* hospitals. Our findings suggest that surgeon performance is not fully portable across hospitals (i.e., some portion of performance is firm specific). Further, we provide preliminary evidence suggesting that this result may be driven by the familiarity that a surgeon develops with the assets of a given organization.

*Key words*: learning; firm-specific performance; freelance workers; productivity; health care
*History*: Accepted by John Boudreau, organizational behavior, performance, strategy, and design; received April 28, 2004. This paper was with the authors 3 months for 1 revision.

## 1. Introduction

Across many sectors of the economy, firms rely on freelance employees or independent contractors for the provision of certain services. These freelancers provide services to multiple firms or move frequently from one firm to the next on short-term assignments. While some markets, such as those for actors, journalists, and consultants, have long depended on such contractual arrangements, a trend toward freelancing is also apparent in other skilled professions including medicine, engineering, and even senior management. Reduction of fixed costs and greater flexibility are generally cited as the chief benefits of these freelance relationships for firms (Davis-Blake and Uzzi 1993, Houseman 2001).[1]

Using highly skilled freelancers to provide services that are critical to an organization's competitive advantage, however, would seem to be at odds with resource-based views of strategy. The ability of a freelancer to serve multiple firms simultaneously, or switch rapidly across firms, implies that no single firm could sustain a competitive advantage from its access to that individual. Rents would either be appropriated by the freelancer (via his or her compensation) or—if the freelancer provides comparable services across firms—competed away in the product market.[2] Theories focusing on the complementarity of inputs (e.g., Milgrom and Roberts 1990, Rivkin 2000, Ahuja 2003), however, offer a mechanism by which firms might appropriate rents from the use of freelancers. Specifically, in the presence of specialized, complementary assets, a freelancer's performance may well differ across organizations. To the extent that a freelancer cannot simply replicate his or her performance across organizations, such differences could provide the basis for an appropriable rent stream for the firm.

The above possibility suggests a testable proposition—to the extent that a freelancer's human capital

[1] Mayer and Nickerson (2005) provide empirical evidence that the benefits of using contractors is reduced in the presence of contracting difficulties related to expropriation, measurement costs, or interdependence. They also provide an extensive review of the literature on the effect of using contract workers rather than employees for outcomes such as innovation (Matusik and Hill 1998, Storey et al. 2002) and subjective evaluations of performance (Jarmon et al. 1998).

[2] We note that the inability of a firm to appropriate the rents created by an individual worker is not unique to freelancers. Worker mobility may also prevent firms from appropriating the rents created by their full-time employees (Peteraf 1993).

is firm specific, that individual's performance should vary across organizations. This paper explores empirically two questions along these lines. First, to what degree is the performance of freelancers firm specific? Second, to the degree that firm-specific performance exists, what are the mechanisms that might explain it? We note that the freelancers to whom our study is most relevant are those who must interact regularly with other assets of an organization (e.g., other workers, capital). As such, our findings may generalize more easily to freelance consultants—who typically work on projects with individuals who are employed by client firms—than to journalists who often complete projects independently.

Beyond those studies mentioned above, the prior literature relevant to our analysis emerges from three broad streams, each of which differs from our approach in some respect. The first stream—from the economics literature—discusses the theoretical arguments that create the expectation of firm-specific performance for individual workers (Becker 1962, Jovanovic 1979). This literature has generally focused on the implications of firm-specific capital for market outcomes such as the movement of workers between firms (Parsons 1972, Jovanovic 1979), rather than considering evidence for whether the actual *performance* of workers is firm specific. A second stream of literature focuses on the impact of free agency (i.e., the ability to enter contracts with multiple firms) on market outcomes such as compensation (Kahn 1993, MacDonald and Reynolds 1994), contract duration (Kahn 1993), and the allocation of workers across firms (Cymrot and Dunlevy 1987, Krautmann and Oppenheimer 1994). To the extent that they address worker performance, papers in this category tend to assume that a worker's marginal productivity is constant across firms (i.e., that performance is not firm specific).

The final stream of prior work—and that most closely related to our approach—are those studies that have examined the performance of workers across firms or organizational contexts (Long and McGinnis 1981, Allison and Long 1990, Almeida and Kogut 1999, Groysberg 2001, Rosenkopf and Almeida 2003, Song et al. 2003). These studies take advantage of the movement of workers from one firm (or setting) to another over time to examine the degree to which knowledge or performance is transferable across settings. Examining the performance of a given worker who switches firms raises certain estimation issues, which many of these studies make careful efforts to address. For example, because a worker is employed by only one firm at a given time, it is difficult to separate the effects of changes in organizational context from other changes that might be occurring for the worker over time (such as changes in level of skill or type of tasks performed).

Our approach serves as a complement to these studies by examining the performance of freelance workers *performing the same task* in multiple firms *at roughly the same point in time*. Specifically, we consider the performance of cardiac surgeons across multiple hospitals using data from every patient receiving coronary artery bypass graft (CABG) surgery in Pennsylvania during 1994 and 1995. For the most part, these surgeons are not employed by specific hospitals but rather have contractual relationships with multiple facilities. In this setting, we can examine a given worker across several firms at a single time, thereby avoiding the need to control for factors, such as increased experience or changes in required job tasks, that may also affect a worker's performance as he or she moves sequentially from one firm to another. In addition, we observe workers who perform the same task at different organizations, thereby reducing concerns about differences in the scope of a worker's duties across firms.

Our study also benefits from a clear mechanism through which firm-specific performance—if present—can be observed. We take advantage of the documented "volume-outcome effect" for CABG, which suggests that surgical outcomes improve as a surgeon or hospital increases its recent or cumulative volume of procedures.[3] More generally, this effect is captured in the concept of the "learning curve" (or "experience curve") that has been observed both at the level of individual workers (e.g., Newell and Rosenbloom 1981, Delaney et al. 1998) and organizations (e.g., Argote and Epple 1990, Adler and Clark 1991). We ask whether a surgeon's performance *at a particular hospital* benefits more from a given amount of recent volume at that hospital than from the same amount of recent volume at another hospital. By observing a surgeon's outcomes across several hospitals within a given year, we are able to examine whether the performance of surgeons is firm specific. We conclude by offering a preliminary investigation of the mechanisms that may generate the effects we observe.

[3] Several studies have found that performance improves with increases in the number of procedures performed by a given surgeon (Hannan et al. 1989, 1991), while others have identified a similar relationship with respect to volume at the hospital level (Luft et al. 1979, Showstack et al. 1987, Pisano et al. 2001). Reagans et al. (2005) find that procedure time declines with volume at the individual, team, and organizational levels. With the exception of Pisano et al. (2001), these studies have used volume during a recent time period (e.g., within the prior year), rather than cumulative volume, as the measure of experience. While learning is traditionally viewed as a function of cumulative experience, some studies have suggested that experience may depreciate over time (Argote et al. 1990, Benkard 2000), thereby highlighting the importance of considering *recent* volume as a predictor of performance.

## 2. The Determinants of Freelancer Performance

Broadly speaking, there are three types of factors that can impact the performance of a worker within a particular firm. These are: (1) worker effects that are independent of the firm; (2) firm effects that are independent of the worker; and (3) effects that are specific to worker-firm combinations. While not the focus of our study, the theoretical explanations for the first two categories of effects have received significant attention in the literature and help motivate our analysis of worker-firm effects. Below we briefly consider worker effects and firm effects in a single discussion and then proceed to motivate our analysis of firm-specific performance.

### Worker and Organizational Determinants of Individual Performance

A worker's performance in any given setting may simply reflect his or her "endowed" skills. In some instances, these skills are innate. For example, athletes or surgeons with superior hand-eye coordination may perform better than less-able individuals with otherwise identical levels of training and experience. Alternatively, worker performance may improve due to changes in skills over time due to factors such as education and training (Becker 1962) or, as mentioned above, experience.[4]

An individual worker's performance can also be influenced by the organization in which he or she works. For instance, organizational factors such as capital equipment, technology, organizational processes, culture, management, and human resource policies may exert an effect on worker performance.[5] These differences in firm-level factors should influence the average performance of workers at that organization.

### Firm-Specific Performance

The theoretical justifications for worker effects and firm effects are quite straightforward. Workers with better skills will, on average, perform better than those with worse skills (holding organizational factors constant), and organizations with better resources should improve the level of worker performance (holding worker skill constant). The question we wish to explore is whether specific worker-firm combinations might yield higher performance than others. Why might a given worker's performance be better at one organization versus another, even after taking into account organizational factors that might influence average performance? Or, posed from the perspective of the firm, why might the performance of two similarly skilled workers vary if each is provided the same firm-level resources?

One category of explanations for firm-specific performance revolves around the potential complementarity between a worker and the human, physical, or organizational assets held by a given firm. For example, a worker that has developed deep skills with respect to information technology may have significantly better performance, on average, than a worker without such skills. Nevertheless, those skills may be less valuable if the firm within which the tech-savvy individual works has not made significant investments in technology as well. A similar argument would hold for a firm that has invested heavily in technology, but has not employed workers capable of applying that technology to its most productive uses.

Another example of complementarity—and one that is particularly relevant to this paper—is the importance of familiarity between members of a team or organization in settings where much of the knowledge required for effective performance is tacit (Polanyi 1966) rather than explicit. This concept of familiarity has been described in the literature using various terms. Chillemi and Gui (1997) refer to the notion "team human capital"—a nonmaterial asset that is derived from customs developed by the members of a team. Similarly, Mailath and Postlewaite (1990) note that a firm includes a "network of workers." With respect to these workers, they note: "They know whom to contact about particular problems that may arise and they know the strengths and weaknesses of their co-workers. A worker has substantial network specific human capital, which is of no value outside of this network" (pp. 369–370).

The importance of familiarity has been observed in several settings. For example, Pisano et al. (2001) note that well-developed surgical teams are often capable of performing procedures with minimal verbal communication between members. In their analysis of flight decks on aircraft carriers, Weick and Roberts (1993) suggest that familiarity is most beneficial not because it leads to habit formation, but rather because it provides team members with a common base of experience that fosters future learning. Katz (1982) and Berman et al. (2002) also find positive effects of familiarity with respect to the performance of teams in industrial research and professional basketball, respectively, though they suggest that these relationships may not be linear at high levels of familiarity.

As suggested above, the beneficial effects of accumulated tacit knowledge within one organization may not be fully transferable to another. Consider a

[4] See Delaney et al. (1998) for a discussion of the literature on individual learning.

[5] Hayes and Wheelwright (1984) and Hayes and Clark (1986) provide general descriptions of these factors. Ichniowski et al. (1997) discuss the impact on productivity of firm-level policies with respect to human resource management.

surgeon who splits her time between two hospitals. As a result of doing more of her operations at one hospital, she develops a deep familiarity with the tacit aspects of the surgical teams at that facility. For instance, she might learn that the surgical nurses at the hospital generally do not speak up about possible problems during an operation, but that, if they do raise an alarm, it signals a very serious problem. The surgeon might also become familiar with the various practice habits of the anesthesiologist with whom she operates or she might gain a sense of which residents she can call upon for quick and reliable information about one of her recovering patients. One can imagine that such familiarity helps the surgeon perform better. In contrast, when that same surgeon ventures across town to operate at a different hospital where she does relatively few cases, the benefits of familiarity at the first hospital may not carry over to the second. Further, her lack of familiarity with the nuances of the environment may detract from her performance at the second hospital.

A second group of explanations might attribute firm-specific individual performance to differences in the influence that an individual wields within particular organizations (Milgrom and Roberts 1988). With respect to our empirical setting, a cardiac surgeon who performs a large number of procedures at a given hospital may achieve better outcomes at that hospital than at other facilities, not simply because of greater familiarity with the staff in the operating room, but also because her high procedural volume makes her someone who can command preferential access—relative to other cardiac surgeons—to the hospital's resources. For example, to the extent that a given hospital has several anesthesiologists or nursing teams of varied quality, an influential surgeon may be able to demand access to the top performers in staffing their surgical teams. There are cases where hospitals have actually dedicated operating rooms to renowned surgeons with high operating volumes. In turn, such preferential access to resources may enable the surgeon to achieve better surgical outcomes. While our primary objective in this study is to determine the degree to which surgeon performance is firm specific, we provide some discussion later in this paper of our efforts to distinguish between familiarity and influence as explanations for firm-specific performance.

## 3. Setting and Data

### Description of CABG

Developed in the late 1960s, CABG is an invasive surgical procedure that involves taking a section of vein (from the leg) or artery (from the chest) and grafting it to create a bypass of blockage in the coronary artery. It requires opening the patient's chest and relies on a heart-lung bypass machine to perform the functions of the heart during the grafting process. For several reasons, CABG represents an instructive setting in which to analyze the firm-specific performance of freelancers. First, cardiac surgeons are archetypal freelancers. They are highly trained, typically receiving up to seven years of residency and fellowship following their four years of medical school. The fact that many states publicly report CABG outcomes by surgeon suggests that these individuals are perceived as being integral to the quality of cardiac care.

Second, while a surgeon is clearly a critical worker, he or she is only one member of a larger surgical team that includes anesthesiologists, nurses, perfusionists, and other technicians. Unlike most surgeons, the other members of a surgical team are typically employed by a hospital. In addition to employing key team members, hospitals also provide a wide range of other organizational assets (e.g., operating room, equipment, marketing, and managerial expertise) that may have an impact on the quality of CABG outcomes regardless of the skill of an individual surgeon.

Third, there exists broad agreement concerning the appropriate measure of performance with respect to CABG—risk-adjusted mortality. Much of the clinical literature on CABG uses some measure of in-hospital or long-term (e.g., several months to several years) mortality following the procedure. This outcome is easily and accurately measured, and it is characterized by enough variation across doctors and hospitals to make it a meaningful dimension for performance evaluation.

Finally, CABG patients account for a significant portion of hospital revenues. In 2000, over 27,000 CABG procedures were performed in Pennsylvania and the average hospital charge for each admission involving CABG was roughly $59,900 (Pennsylvania Health Care Cost Containment Council 2002). While charges represent the "list prices" for hospitals, the total revenue that Pennsylvania hospitals derived from CABG admissions—assuming that hospitals collected anywhere from 50% to 70% of charges—would range from $800 million to $1.1 billion. Scaling these figures up to the roughly 355,000 CABG patients in the United States during 1999 (American Heart Association 2001) suggests nationwide hospital revenues of between $10.5 billion and $14.5 billion for CABG patients alone. This range represents between 3% and 4% of total patient revenue ($342 billion in 2000) for all hospitals in the United States (Health Forum 2002).

In addition to the benefits mentioned above, studying CABG *within Pennsylvania* allows clean identification of firm-specific performance. As noted earlier, there is a well-documented relationship between annual procedure volume—for both hospitals and surgeons—and mortality outcomes for CABG. Nevertheless, prior studies have not considered the importance of a surgeon's hospital-specific volume. Given the prevalence of surgeons who split their time across hospitals in Pennsylvania, we are able to separate the impact of a surgeon's hospital-specific volume from that of his or her volume at other hospitals.

### Data

The data for this analysis are from the Pennsylvania Health Care Cost Containment Council (PHC4) and include patient-level records for every individual receiving CABG at a hospital in Pennsylvania in 1994 or 1995. These data cover 38,577 procedures performed by 203 surgeons operating at 43 hospitals.[6] In addition to identifying the hospital and surgeon for each procedure, PHC4 also provides a broad range of demographic and clinical information for each patient. This information includes patient age, gender, illness severity upon hospital admission, and a series of variables indicating the presence of particular comorbidities such as kidney failure, heart failure, and acute myocardial infarction (i.e., heart attack).

### Calculation of the Risk-Adjusted Mortality Rate

One method for comparing quality across surgeons or hospitals is by examining differences in their rates of in-hospital patient mortality. Given heterogeneity in the severity of patients' preoperative conditions, however, raw (i.e., observed) mortality rates represent potentially-biased measures of the quality of surgeon performance. In particular, higher quality surgeons may attract patients with more severe forms of coronary disease, and these individuals are more likely to die in the hospital independent of provider quality.

To mitigate this bias, PHC4 performs logistic regression on patient-level observations for nearly every CABG procedure conducted at a hospital in Pennsylvania during 1994 and 1995.[7] This regression controls for several patient characteristics or existing clinical conditions (e.g., age, gender, complicated hypertension, heart failure, heart attack, kidney failure, cardiogenic shock, and others) that could affect a patient's underlying probability of dying in the hospital.[8] The dependent variable in this regression, $\mathrm{MORT}_{i,s,h}$, is an indicator equal to one if patient $i$—who received CABG from surgeon $s$ at hospital $h$—died in the hospital, and zero otherwise. PCH4 calculates the predicted probability of death for each patient as the fitted value for that individual obtained from the logistic regression. The form of this logistic regression is as follows:

$$\ln\left(\frac{\Pr(\mathrm{MORT}_{i,s,h}=1 \mid x_i)}{1-\Pr(\mathrm{MORT}_{i,s,h}=1 \mid x_i)}\right)=\alpha_0+\alpha_1\cdot X_i+\varepsilon_{i,s,h}, \quad (1)$$

where $X_i$ is a vector of patient-level clinical variables.

To calculate the risk-adjusted mortality rate for a given hospital ($\mathrm{RAMR}_h$), we average the *predicted* probability of mortality from (1) across all patients at that hospital to create the expected mortality rate ($\mathrm{EMR}_h$) for that facility during a given time period. Similarly, the observed, or actual, mortality rate ($\mathrm{OMR}_h$) is the total number of deaths at hospital $h$ divided by the total number of procedures at $h$ during the same time period. $\mathrm{RAMR}_h$ is then calculated as follows:

$$\mathrm{RAMR}_h=(\mathrm{OMR}_h/\mathrm{EMR}_h)*\mathrm{OMR}_{\mathrm{PA}}, \quad (2)$$

where $\mathrm{OMR}_{\mathrm{PA}}$ is the average observed mortality rate for the entire state of Pennsylvania over the time period. This multiplication serves to normalize the ratio of observed-to-expected mortality to the statewide average for CABG mortality.[9]

### Splitters and Nonsplitters

There are several potential explanations for why surgeons split their time across hospitals. For example, surgeons with strong reputations may draw patients from a relatively broad geography and may offer multiple hospital options to ensure patient convenience. In a related vein, many surgeons are members of multiphysician group practices. To the extent that surgeons provide coverage for their colleagues—who tend to practice at different hospitals—one would expect to see some degree of splitting activity. Splitting patterns may have also resulted from the increase in mergers and other affiliations between hospitals in the 1990s. To the extent that these transactions linked multiple hospitals, each of which had its own CABG program, one might expect to see surgeons splitting their time across these facilities as part of a "systemwide" CABG program. In building up these programs, hospitals also tried to recruit cardiac surgeons

[6] Due to our use of lagged independent variables, CABG procedures performed during the first quarter of 1994 (and, in some cases, the first two quarters of 1994) are excluded from our base regressions in Tables 3 and 4. As a result, the sample sizes for these regressions range from 28,173 to 34,173 cases.

[7] There are several criteria used to exclude a patient from the analysis, and they are detailed in Pennsylvania Health Care Cost Containment Council (1998). The final sample for PHC4's logistic regression included 38,577 CABG cases for the 1994–1995 period.

[8] A full list of the covariates included in this regression, as well as the resulting coefficent estimates, can be found in Pennsylvania Health Care Cost Containment Council (1998).

[9] Later in the paper, we also calculate similar risk-adjusted rates, $\mathrm{RAMR}_s$ and $\mathrm{RAMR}_{s,h}$, at the surgeon and surgeon-hospital levels, respectively.

**Table 1 Comparison of Splitters and Nonsplitters**

| | Splitters ($n = 357$) | | Nonsplitters ($n = 996$) | |
|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation |
| CABG cases/surgeon per calendar quarter (1994 and 1995 combined) | 32.9 | 17.7 | 27.0 | 17.0*** |
| Risk-adjusted mortality rate (1994 and 1995 combined) (%) | 3.35 | 3.42 | 3.01 | 4.24 |

*Notes.* *** denotes that the mean value for splitters and nonsplitters are significantly different at the 1% level. The level of observation is a surgeon for a given calendar quarter. A surgeon is a "splitter" for a given quarter if he or she performed at least one CABG procedure at each of two or more hospitals in Pennsylvania during that quarter.

from other facilities. These recruitment efforts often resulted in surgeons splitting their time across organizations as they tested out a new facility. Finally, some surgeons may split their time because capacity at their hospital of choice is constrained and they need to look elsewhere to find operating room time. The PHC4 data does not provide information as to which, if any, of the above factors explain splitting behavior. As such, this paper focuses on the *effects*, rather than the *causes*, of this activity.

Table 1 provides simple comparisons of volumes and performance for "splitters" and "nonsplitters." For this initial analysis, we consider a surgeon to be a splitter during a particular quarter of the calendar year if he or she performed at least one procedure at each of two or more hospitals during that calendar quarter. Based on this criterion, roughly 30% of the cardiac surgeons in the sample were splitters during the average quarter during the sample period. Further, a surgeon who is a splitter in one quarter tends to be a splitter in the subsequent quarter; the correlation between being a splitter in the prior quarter and the current quarter is 0.82.

Each observation in Table 1 represents a surgeon in a given three-month quarter during the 1994–1995 period. The average splitter performed 5.9 (22%) more procedures per quarter than the average nonsplitter, and this difference is significant at the 1% level. We note that the lower CABG volume of nonsplitters does not appear to be due to their substituting non-CABG surgeries, such as valve replacements, for CABG procedures. In fact, the average number of non-CABG surgeries is very small—roughly one per quarter—for both groups. Though the RAMR for splitters is slightly higher than that for nonsplitters, this difference is not significant at conventional levels. The online supplement to this paper (http://mansci.pubs.informs.org/ecompanion.html) provides illustrative data on the performance and case volumes for a few splitters in the sample.

## 4. Empirical Strategy and Predictions

A common empirical approach for measuring the firm specificity of performance is to examine individual workers who move across multiple teams or organizations (Long and McGinnis 1981, Allison and Long 1990, Almeida and Kogut 1999, Groysberg 2001, Rosenkopf and Almeida 2003, Song et al. 2003). This approach takes advantage of an individual's change of team or employer over time. Nevertheless, the fact that an individual is typically employed by a single firm—or involved with a single team—at any given point in time creates difficulty in distinguishing the effect of switching teams or firms from other time-varying factors such as general (i.e., not firm specific) learning by doing.[10]

The empirical setting for our study allows us to abstract from the confounding effect noted above. Specifically, we observe highly skilled individuals who, as freelancers, split their time across multiple organizations *roughly simultaneously*.[11] The movement of surgeons across hospitals during this period enables us to obtain separate estimates for each of the three effects—surgeon specific, hospital specific, and surgeon-hospital specific—described above without having to worry about changes in a surgeon's level of general experience or skill over time.

**Impact of Total Surgeon Volume**

We take advantage of the identified relationship between volume of activity and performance in cardiac surgery to test several hypotheses concerning surgeon performance. Prior to testing our main hypothesis—that some portion of surgeon performance is firm specific—we first need to establish the basic relationship between a surgeon's total (i.e., across all hospitals) volume of procedures and his or her overall performance. This analysis is used to confirm that our data illustrates the volume-outcome effects identified in prior studies of CABG procedures using data from other states, such as New York

[10] Some studies control for learning by including a measure of a worker's overall experience. In settings where individuals learn at different rates, however, such measures may not be able to control fully for learning over time. As an analogue to differences in rates of learning across individuals, several studies discuss differences in the rate of learning across organizations (Argote and Epple 1990, Pisano et al. 2001).

[11] By "roughly simultaneously," we mean within the course of a relatively short period of time, such as one week or one month. For the purpose of protecting patient confidentiality, the Pennsylvania data only allows one to identify the calendar quarter—not the month or day—in which individual procedures were performed. Nevertheless, most surgeons who split their time across hospitals appear to do so evenly across the four quarters of the year. We therefore assume that these surgeons likely split their time evenly within time periods shorter than calendar quarters (e.g., months or weeks).

(Hannan et al. 1989, 1991). Our basic logistic regression takes the following form:

$$\ln\left(\frac{\Pr(\mathrm{MORT}_{i,s,h}=1\mid z_i)}{1-\Pr(\mathrm{MORT}_{i,s,h}=1\mid z_i)}\right) = \gamma_0+\gamma_1\cdot \mathrm{TOTCASE}_{s,q-1}+\gamma_2\cdot \mathrm{RAMR}_{s,q-1} + \gamma_3\cdot \mathrm{RAMR}_{h,q-1}+\gamma_4\cdot Z_i+\varepsilon_{i,s,h}, \tag{3}$$

where observations are at the level of the individual patient $i$, who receives treatment from surgeon $s$ at hospital $h$. $\mathrm{MORT}_{i,s,h}$ is the indicator for patient mortality defined in (1). To reduce concerns regarding the direction of causality, we lag all of the other independent variables by one quarter. Thus, $\mathrm{TOTCASE}_{s,q-1}$ is the number of CABG cases performed by surgeon $s$ at all Pennsylvania hospitals during the prior quarter. To check the robustness of our findings, we repeat this regression replacing the total cases in the prior quarter with the total cases in the current quarter.

To control for the fact that surgeons and hospitals have different underlying levels of quality due to factors that are independent of procedure volume, we include two additional variables. The first, $\mathrm{RAMR}_{s,q-1}$, is the risk-adjusted mortality rate for surgeon $s$ across *all* hospitals in the prior quarter. This variable captures the effect of surgeon quality on performance. Similarly, the second variable, $\mathrm{RAMR}_{h,q-1}$, is the same measure for hospital $h$ across *all* surgeons in the prior quarter; it controls for the impact of hospital quality on outcomes.[12] Finally, $Z_i$ is a vector that includes all of the patient-level clinical variables included in (1) as well as fixed effects for each calendar quarter. To address any lack of independence in the error terms, we cluster the standard errors in (3) by surgeon.

In accordance with the findings of prior studies, we expect the coefficient on TOTCASE[13] to be negative. That is, an increase in a surgeon's procedure volume across all hospitals during the previous quarter should, on average, reduce his or her risk-adjusted mortality rate. Further, we expect the coefficients on both of the lagged mortality rates ($\mathrm{RAMR}_s$ and $\mathrm{RAMR}_h$) to be positive—lower mortality in a prior period should be correlated with lower mortality in the current period.

[12] One potential concern with using one-quarter-lagged quality measures is that they may not capture the true underlying quality of a given surgeon or hospital. As a robustness check, we perform our regressions with six-month-lagged quality measures and with no quality measures. The results of these sensitivity tests are discussed later in the paper.

[13] For notational simplicity, we present the subscripts for variables upon first mention and then only as necessary to differentiate the level or time period for which the variable is reported.

**Impact of Hospital-Specific Surgeon Volume**

Next, we test the hypothesis that some portion of surgeon performance is firm specific. Again, we take advantage of the positive relationship between the procedural volume of surgeons and the outcomes they achieve. As freelancers, many of the surgeons in our study perform procedures at multiple hospitals within the course of short periods of time (e.g., one week).

For example, a given surgeon may perform procedures at Hospital A on Mondays and Tuesdays and at Hospital B on Wednesdays, Thursdays, and Fridays. The movement of this surgeon between hospitals thus allows us to consider not only the degree to which her volume of procedures *at Hospital A* improves her performance at Hospital A, but, more importantly, the extent to which her volume of procedures *at Hospital B* improves her performance at Hospital A. To the extent that a surgeon's performance at Hospital A is improved more by volume at Hospital A than by the same volume at Hospital B, one can consider surgeon performance to be firm specific. An alternative statement of this prediction is that, if the performance of surgeons were not at all firm specific, then one would expect volume at Hospitals A and B to have identical effects on the surgeon's outcomes at Hospital A (i.e., the benefits of surgeon experience obtained at one hospital would be fully portable to the other).

To test for firm specificity in surgeon performance, we estimate a variant of (3) in which a surgeon's total volume of procedures in the prior quarter is segmented into his or her volume at hospital $h$ and that at all other hospitals. This specification appears below:

$$\ln\left(\frac{\Pr(\mathrm{MORT}_{i,s,h}=1\mid z_i)}{1-\Pr(\mathrm{MORT}_{i,s,h}=1\mid z_i)}\right) = \beta_0+\beta_1\cdot \mathrm{HOSPCASE}_{s,h,q-1} + \beta_2\cdot \mathrm{OTHCASE}_{s,h,q-1}+\beta_3\cdot \mathrm{RAMR}_{s,q-1} + \beta_4\cdot \mathrm{RAMR}_{h,q-1}+\beta_5\cdot Z_i+\mu_{i,s,h}. \tag{4}$$

Again, the level of observation is the individual patient, $i$, who receives treatment from surgeon $s$ at hospital $h$. $\mathrm{HOSPCASE}_{s,h,q-1}$ is the number of CABG cases performed by surgeon $s$ at hospital $h$ in the prior quarter, and $\mathrm{OTHCASE}_{s,h,q-1}$ is the number of cases performed by that surgeon at all Pennsylvania hospitals other than $h$ in the prior quarter.[14]

[14] Based on this patient-level structure, it is clear that one need not arbitrarily assign hospitals to HOSPCASE and OTHCASE. For example, assume that a surgeon performs 100 procedures in a given calendar quarter, 90 at Hospital A and 10 at Hospital B. For that quarter, he would have 100 observations in the data. For 90 of them, HOSPCASE would be the number of procedures he per-

Analogous to our predictions with respect to (3), we expect the coefficient on HOSPCASE to be negative. We would expect the coefficient on OTHCASE to be smaller in absolute magnitude than that for HOSPCASE and either negative or insignificantly different from zero. A strong test for firm specificity in performance requires that the coefficient on OTHCASE is not only smaller in absolute magnitude than—but also significantly different from—that on HOSPCASE. If this is true, then one could conclude that the only volume that significantly improves a surgeon's performance at hospital $h$ is his volume at hospital $h$, not his volume at other hospitals.

We note that (4) includes cases not only for surgeons who split their time across hospitals, but also those for surgeons who operated exclusively at one facility in the prior quarter. As a test of robustness, we repeat our basic analysis using only those cases performed by surgeons who split their time across hospitals during the prior quarter.

**The Roles of Familiarity and Surgeon Influence**

As noted in §2, firm specificity in surgeon performance may result from two factors—familiarity and influence. Empirically separating these two explanations, however, is not a trivial task. The main difficulty is deciding upon an appropriate proxy for a surgeon's influence (relative to other cardiac surgeons) at a particular hospital. To address this issue, we examine two potential proxies for influence: a surgeon's share of total CABG volume at a hospital and a surgeon's share of the academic citation activity at a hospital. We discuss the motivation for each of these proxies below.

Our first proxy for influence is $\text{SHARE}_{s,h,q-1}$, the CABG volume of surgeon $s$ at hospital $h$ as a share of hospital $h$'s total CABG volume in the prior quarter. To the extent that the coefficient on HOSPCASE is negative in (4), one cannot determine whether outcomes are associated with a surgeon's *absolute* volume at a particular hospital—a measure of familiarity—or his volume *relative* to other cardiac surgeons—a potential measure of influence. We note that SHARE captures the *within-specialty* influence of a given surgeon. That is, it measures a surgeon's influence among cardiac surgeons, but does not capture the influence of cardiac surgery relative to other programs at a particular hospital. We use this measure of within-specialty influence for two reasons. First, relative to other departments within a hospital, cardiac surgery tends to be influential due to its high level of profit (Huckman 2003). As a result, cardiac surgery is a service that is highly valued by nearly every hospital that provides it. Second, due to the specialized assets and training required for cardiac surgery, hospitals that offer the service usually have dedicated operating rooms and technical staff for heart procedures. To the extent that cardiac surgeons compete for resources in terms of operating room scheduling or team members, they thus compete against one another.

One limitation of the SHARE proxy is that it may be correlated with our measure of familiarity (HOSPCASE) and, as a result, may not enable us to separate the two explanations for firm-specific performance. Further, it is not clear that the influence of a surgeon relative to his or her colleagues is adequately captured by volume-based measures. For example, surgeons are often part of multiphysician groups comprised of senior and junior partners. While the latter may perform the bulk of a group's procedures, the former may have many years of experience or high status within the hospital community. As such, one might argue that low-volume senior partners may still wield significant influence within a given hospital. Our second proxy addresses this possibility.

Our second measure of influence is $\text{CITATIONS}_{s,h,q}$. It is the number of citations to a surgeon's academic publications as a percentage of the total CABG-related citations for all cardiac surgeons operating at hospital $h$ in quarter $q$. To understand this measure as a proxy for influence, we note that other studies have used bibliometric measures based on patent or academic citations to capture the value of innovations and, in turn, the status of innovators within scientific communities (e.g., Trajtenberg 1990, Podolny and Stuart 1995, Huckman 2003). Given that a large portion of the hospitals offering CABG are listed as teaching hospitals,[15] one can assume that citations in the academic literature may capture the influence of an individual surgeon relative to his or her colleagues at a given hospital.

We calculate CITATIONS as follows:

$$\text{CITATIONS}_{s,h,q} = \frac{C_s}{\sum_{s=1}^{S_{h,q}} C_s}, \tag{5}$$

where the numerator is the total number of CABG-related citations for publications by surgeon $s$ while

formed at Hospital A in the prior quarter and OTHCASE would be the number that he performed at Hospital B in the prior quarter. For the remaining 10 observations, the values of HOSPCASE and OTHCASE would be reversed.

[15] Roughly 90% of the hospitals in Pennsylvania that offered CABG in 1994 or 1995 reported a medical school affiliation to the American Medical Association and approximately 50% were members of the Council of Teaching Hospitals of the Association of American Medical Colleges. Finally, over 85% of the CABG procedures in Pennsylvania were performed at hospitals that had at least one surgeon with academic citations between 1970 and 1993.

the denominator is the total number of citations for CABG-related articles by all $S_{h,q}$ cardiac surgeons operating at hospital $h$ during quarter $q$. The online supplement to this paper contains additional details on the construction of this variable.

Adding each of these proxies into our estimates of (4) changes the interpretation of the coefficient on one of our key variables of interest—HOSPCASE. Specifically, the coefficient on HOSPCASE can now be viewed as a measure of how performance depends on a surgeon's absolute volume (i.e., familiarity) at a specific hospital after controlling for his or her influence relative to other cardiac surgeons. To the extent that greater influence is responsible for the relationship between firm-specific volume and performance, we would expect the coefficient on each of our share proxies to be negative and significant. Further, we would expect the gap between the coefficients on our measures of absolute volume—HOSPCASE and OTHCASE—to be smaller than in our estimates of (4). To the extent that familiarity drives the relationship between firm-specific volume and performance, we would predict the direction—and relative significance—of HOSPCASE and OTHCASE to be only slightly, if at all, affected by the inclusion of our influence proxies.

**Table 2** **Descriptive Statistics for Key Variables in Logistic Regressions**

| | Observations | Mean | Standard deviation |
|---|---|---|---|
| $MORT_{i,s,h}$: Did CABG patient die in hospital? | 34,173 | 3.1% | 17.3% |
| $TOTCASE_{s,q-1}$: Total surgeon cases (prior 3 months) | 34,173 | 36.8 | 17.4 |
| $HOSPCASE_{s,h,q-1}$: Surgeon cases at hospital (prior 3 months) | 34,173 | 32.3 | 17.7 |
| $OTHCASE_{s,h,q-1}$: Surgeon cases at other hospitals (prior 3 months) | 34,173 | 4.5 | 10.0 |
| $RAMR_{s,q-1}$: Surgeon RAMR (prior 3 months) | 33,610 | 3.1% | 4.5% |
| $RAMR_{h,q-1}$: Hospital RAMR (prior 3 months) | 34,173 | 3.1% | 1.9% |
| $SHARE_{s,h,q-1}$: Surgeon share of total hospital CABG cases (prior 3 months) | 34,173 | 27.0% | 20.0% |
| $CITATIONS_{s,h,q}$: Surgeon share of academic CABG citations at hospital for 1970–1993 | 30,572 | 22.4% | 34.6% |

*Notes.* While there are 38,577 observations in the Pennsylvania database for 1994 and 1995, the largest sample size above (34,173) includes only cases for 1995 and the last three calendar quarters of 1994. Observations for the first quarter of 1994 are lost in calculating lagged volume and mortality measures. The reduction to 33,610 for $RAMR_{s,q-1}$ reflects the fact that a few surgeons may not have performed any procedures in Pennsylvania in the prior quarter (and their risk-adjusted mortality rate for that period must be set to missing). Finally, the reduction to 30,572 observations for $CITATIONS_{s,h,q}$ is due to the fact that some hospitals did not have any surgeons with nonzero citation volume for the relevant period.

# 5. Results and Discussion

**Total Surgeon Volume**

Table 2 provides descriptive statistics for the key variables used in our analyses. Table 3 shows the results for regressions using total surgeon volume (i.e., across *all* hospitals). Column 1 supports our prediction that an increase in a surgeon's total volume in the prior quarter is correlated with a reduction in risk-adjusted mortality. The negative coefficient on TOTCASE, which is significant at the 1% level, is consistent in direction with the hypothesized volume-outcome effect that has been observed at the physician level in prior studies (Hannan et al. 1989, 1991). The rows toward the bottom of the table reveal that an additional one case per quarter is correlated with a decrease of 0.015 percentage points in the probability of mortality. Relative to the predicted probability of mortality evaluated at the means of the independent variables (1.77%), this change represents a decline of 0.85%.

The addition of controls for surgeon quality ($RAMR_{s,q-1}$) and hospital quality ($RAMR_{h,q-1}$) does not substantially affect the magnitude or significance of the coefficient on total surgeon cases (Column 2, Table 3). It is not surprising that the coefficient on $RAMR_{s,q-1}$ is positive and highly significant, as surgeons with worse results across all hospitals in the prior quarter would be expected to have worse results at hospital $h$ in the current period. While the direction of the coefficient on $RAMR_{h,q-1}$ is negative—which runs counter to our expectation—it is not significant at conventional levels. One interpretation of the results for these two controls is that the underlying quality of the surgeon is more important than that of the hospital in determining the quality of future outcomes for the surgeon-hospital pair. Alternatively, the highly insignificant and counterintuitive coefficient on hospital RAMR could simply be due to the correlation (0.37) between the $RAMR_{s,q-1}$ and $RAMR_{h,q-1}$ variables.

We are not concerned about our inability to distinguish between these two explanations empirically. Rather, we are more interested in making sure that we *control* for the effects of underlying quality at both the surgeon and hospital level, and this specification achieves that goal. Nonetheless, we perform a robustness check by re-estimating (3) using surgeon- and hospital-level mortality rates for the prior six months (rather than the prior quarter). We refer to these rates as $RAMR_{s,q-2}$ and $RAMR_{h,q-2}$, respectively. Column 3 shows that the coefficient on our key variable of interest, TOTCASE, remains roughly the same as in our prior estimates when we add these alternate qual-

**Table 3 Logistic Regression with Total Surgeon Volume**

| | Dependent variable: Did CABG patient die in hospital? | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $TOTCASE_{s,q-1}$: Total surgeon cases (prior quarter) | −0.0087*** (0.0023) | −0.0090*** (0.0024) | −0.0099*** (0.0025) | | | |
| $TOTCASE_{s,q}$: Total surgeon cases (current quarter) | | | | −0.0114*** (0.0024) | −0.0113*** (0.0025) | −0.0119*** (0.0027) |
| $RAMR_{s,q-1}$: Surgeon RAMR (prior quarter) | | 0.9293** (0.4826) | | | 0.7976 (0.5039) | |
| $RAMR_{s,q-2}$: Surgeon RAMR (prior 6 months) | | | 0.7809 (0.6306) | | | 0.6787 (0.6636) |
| $RAMR_{h,q-1}$: Hospital RAMR (prior quarter) | | −0.3233 (1.7905) | | | −0.7497 (1.7833) | |
| $RAMR_{h,q-2}$: Hospital RAMR (prior 6 months) | | | 0.7590 (2.6301) | | | 0.2836 (2.6105) |
| Constant | −3.4691*** (1.2687) | −3.7491*** (1.2841) | −3.7928*** (1.4047) | −3.2515*** (1.2559) | −3.5476*** (1.2711) | −3.6835*** (1.3881) |
| Average predicted probability of mortality (evaluated at means of independent variables) (%) | 1.77 | 1.77 | 1.76 | 1.76 | 1.76 | 1.75 |
| Impact of one unit increase in TOTCASE (%) | −0.015 | −0.016 | −0.017 | −0.020 | −0.020 | −0.020 |
| Observations | 34,173 | 33,610 | 28,173 | 34,173 | 33,610 | 28,173 |
| Pseudo $R^2$ | 0.1745 | 0.1753 | 0.1745 | 0.1759 | 0.1765 | 0.1755 |
| Wald chi-squared | 1,941.8*** | 1,933.3*** | 1,727.9*** | 1,906.9*** | 1,914.7*** | 1,708.4*** |

*Notes.* ** and *** denote statistical significance at the 5% and 1% levels, respectively. The following variables are included in the regressions but are not shown in the table: age, age$^2$/100, fixed effects for quarter of calendar year, and indicators for cardiogenic shock, concurrent angioplasty, complicated hypertension, dialysis, female gender, heart failure, and prior CABG or heart valve surgery. Standard errors are heteroskedasticity robust and clustered by surgeon.

ity measures. We note that while the coefficients on both controls are statistically insignificant, they are in the predicted positive direction. Our main finding from this regression, however, is that the coefficient on TOTCASE is negative and that the estimate is not sensitive to the use of three-month quality measures, six-month quality measures, and—as suggested by Column 1—the absence of lagged quality measures.

The analysis in Columns 4 through 6 of Table 3 repeats that in Columns 1 through 3 with the exception that surgeon volume in the *prior* quarter is replaced by surgeon volume in the *current* quarter ($TOTCASE_{s,q}$). We note that the results are similar to those using volume from the prior quarter.

**Hospital-Specific Surgeon Volume**

The results of our test for hospital-specificity in surgeon performance—as described in (4)—appear in Table 4. Column 1 decomposes a surgeon's total case volume into those performed at hospital $h$ (HOSPCASE) and those occurring elsewhere (OTHCASE). Both coefficients are negative, but that on HOSPCASE is significant at 1%, while that on OTHCASE is much smaller in magnitude and insignificant at conventional levels. The marginal effect for HOSPCASE is slightly larger than, although similar in magnitude, to that on TOTCASE in Table 3—an increase of one case per quarter is associated with a 0.018 percentage point decline in the mortality rate. For OTHCASE, a similar increase of one case is correlated with a decline of only 0.001 percentage points that, again, is insignificant at conventional levels.

The strongest evidence of firm specificity in performance is that the coefficient on HOSPCASE is not only significantly different from zero, but is also significantly different from the OTHCASE coefficient at the 1% level. This result, which appears at the bottom of Table 4, implies that surgeon's volume at a given hospital affects her outcome at that hospital significantly more than does her volume at other hospitals. The relative magnitude and significance of these key coefficients—as well as the significance of the difference between them—does not change meaningfully as we add controls for surgeon and hospital quality for the prior quarter (Column 2) and prior six months (Column 3).[16] Finally, these results are robust to the use of volume measures for the current

[16] As in our results using total surgeon volume (TOTCASE), the inclusion of various measures of surgeon quality does not meaningfully affect the coefficients on HOSPCASE. As an additional check, we regress the *one-quarter change* HOSPCASE on the *one-quarter change* in a surgeon's risk-adjusted mortality rate at that hospital and fixed effects for the time period. The rationale behind this regression is that if a surgeon experiences a decline in performance at a particular hospital, she may experience a reduction in her level of procedures at that facility. Nonetheless, the coefficient

**Table 4 Logistic Regression with Local Volume and Volume at Other Hospitals**

| | Dependent variable: Did CABG patient die in hospital? | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $HOSPCASE_{s,h,q-1}$: Surgeon cases at hospital (prior quarter) | −0.0103*** (0.0023) | −0.0108*** (0.0024) | −0.0117*** (0.0025) | | | |
| $HOSPCASE_{s,h,q}$: Surgeon cases at hospital (current quarter) | | | | −0.0134*** (0.0023) | −0.0134*** (0.0024) | −0.0143*** (0.0027) |
| $OTHCASE_{s,h,q-1}$: Surgeon cases at other hospitals (prior quarter) | −0.0003 (0.0037) | −0.0008 (0.0037) | −0.0020 (0.0037) | | | |
| $OTHCASE_{s,h,q}$: Surgeon cases at other hospitals (current quarter) | | | | −0.0012 (0.0036) | −0.0011 (0.0037) | −0.0008 (0.0037) |
| $SURGRAMR_{s,q-1}$: Surgeon RAMR (prior quarter) | | 0.8762* (0.4855) | | | 0.7395 (0.5096) | |
| $SURGRAMR_{s,q-2}$: Surgeon RAMR (prior 6 months) | | | 0.7320 (0.6467) | | | 0.6246 (0.6878) |
| $HOSPRAMR_{h,q-1}$: Hospital RAMR (prior quarter) | | −0.4988 (1.7617) | | | −1.0445 (1.7611) | |
| $HOSPRAMR_{h,q-2}$: Hospital RAMR (prior 6 months) | | | 0.3717 (2.5858) | | | −0.4616 (2.5724) |
| Constant | −3.5207*** (1.2689) | −3.7893*** (1.2843) | −3.8474*** (1.4062) | −3.2406*** (1.2529) | −3.5114*** (1.2681) | −3.6542*** (1.3842) |
| Average predicted probability of mortality (evaluated at means of independent variables) (%) | 1.76 | 1.76 | 1.75 | 1.75 | 1.75 | 1.73 |
| Impact of one unit increase in: | | | | | | |
| HOSPCASE (%) | −0.018 | −0.019 | −0.020 | −0.023 | −0.023 | −0.024 |
| OTHCASE (%) | −0.001 | −0.001 | −0.003 | −0.002 | −0.002 | −0.001 |
| Level at which HOSPCASE is significantly different from OTHCASE (%) | 1 | 1 | 1 | 1 | 1 | 1 |
| Observations | 34,173 | 33,610 | 28,173 | 34,173 | 33,610 | 28,173 |
| Pseudo $R^2$ | 0.1754 | 0.1762 | 0.1753 | 0.1771 | 0.1777 | 0.1770 |
| Wald chi-squared | 1,979.7*** | 1,969.0*** | 1,776.0*** | 1,964.3*** | 1,975.5*** | 1,766.0*** |

*Notes.* * and *** denote statistical significance at the 10% and 1% levels, respectively. The following variables are included in the regressions but are not shown in the table: age, age²/100, fixed effects for quarter year, and indicators for cardiogenic shock, concurrent angioplasty, complicated hypertension, dialysis, female gender, heart failure, and prior CABG or heart valve surgery. Standard errors are heteroskedasticity robust and clustered by surgeon.

quarter rather than for the prior quarter (Columns 4 through 6).[17]

As an additional check on the robustness of these findings, we repeat our regressions including only those patients who received CABG from a surgeon who split her time across multiple hospitals in the prior quarter. The purpose of this check is to make sure that the coefficient on HOSPCASE in Table 4 is not simply driven by the *total* volume of cases performed by surgeons who did *not* split their time across facilities. Table 5 presents our results for splitters only. Despite a significantly smaller sample of only 10,150 cases (versus more than 33,000 for the initial regressions), the magnitude and significance of the HOSPCASE and OTHCASE coefficients remain similar to those reported in Table 4. The main difference is that HOSPCASE and OTHCASE coefficients are significantly different from each at the 6% level of significance (using three-month lagged quality measures) or the 11% level (using six-month lagged quality). Given the similarity of the HOSPCASE and OTHCASE estimates to those in Table 4, it is possible—although not certain—that the lower significance levels for these models may be due to the dramatic reduction in sample size for these regressions. Despite the smaller sample, the evidence of firm specificity in surgeon performance remains strong.

on the change in mortality in this regression is very low (−0.30) and is insignificant.

[17] To examine robustness, we estimate two additional models. In the first model, lagged hospital quality ($RAMR_{h,q-1}$) is replaced by hospital fixed effects. This specification accounts for the possibility that mortality outcomes may be affected by fixed, hospital-level variables that are not fully captured by lagged quality. In the second model, we replace the HOSPCASE and OTHCASE with their natural logarithms; to account for zero values, we add one to each value before taking the logarithm. We find our results to be robust to these alternate specifications.

**Table 5 Logistic Regression with Local Volume and Volume at Other Hospitals (Splitters Only)**

| | Dependent variable: Did CABG patient die in hospital? | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| $\text{HOSPCASE}_{s,h,q-1}$: Surgeon cases at hospital (prior quarter) | −0.0112*** (0.0044) | −0.0114*** (0.0045) | −0.0125*** (0.0051) |
| $\text{OTHCASE}_{s,h,q-1}$: Surgeon cases at other hospitals (prior quarter) | −0.0007 (0.0052) | −0.0007 (0.0053) | −0.0028 (0.0053) |
| $\text{SURGRAMR}_{s,q-1}$: Surgeon RAMR (prior quarter) | | 0.8809 (1.3484) | |
| $\text{SURGRAMR}_{s,q-2}$: Surgeon RAMR (prior 6 months) | | | 1.3299 (2.2751) |
| $\text{HOSPRAMR}_{h,q-1}$: Hospital RAMR (prior quarter) | | −1.9355 (2.2016) | |
| $\text{HOSPRAMR}_{h,q-2}$: Hospital RAMR (prior 6 months) | | | −3.7666 (3.5599) |
| Constant | −4.3770** (2.0318) | −4.3225** (2.0165) | −4.0230* (2.2421) |
| Average predicted probability of mortality (evaluated at means of independent variables) (%) | 1.88 | 1.88 | 1.86 |
| Impact of one unit increase in: | | | |
| HOSPCASE (%) | −0.021 | −0.021 | −0.023 |
| OTHCASE (%) | −0.001 | −0.001 | −0.005 |
| Level at which HOSPCASE is significantly different from OTHCASE (%) | 6 | 6 | 11 |
| Splitters only? | Yes | Yes | Yes |
| Observations | 10,150 | 10,150 | 8,620 |
| Pseudo $R^2$ | 0.1755 | 0.1757 | 0.1844 |
| Wald chi-squared | 715.9*** | 736.9*** | 714.7*** |

*Notes.* *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The following variables are included in the regressions but are not shown in the table: age, age$^2$/100, fixed effects for quarter year, and indicators for cardiogenic shock, concurrent angioplasty, complicated hypertension, dialysis, female gender, heart failure, and prior CABG or heart valve surgery. Standard errors are heteroskedasticity robust and clustered by surgeon.

**Familiarity vs. Surgeon Influence**

The results in Tables 4 and 5 provide evidence of firm specificity in surgeon performance, but they do not explain the degree to which that hospital specificity is driven by familiarity or surgeon influence. To begin to understand these mechanisms, we add proxies for surgeon influence to our basic model and report the resulting estimates in Table 6. In Column 1, SHARE—surgeon $s'$ share of all CABG procedures at hospital $h$ in the prior quarter—enters as a continuous variable. In Column 2, $\text{SHARE}_{s,h,q-1}$ is replaced by $\text{HIGH\_SHARE}_{s,h,q-1}$—an indicator equal to one if surgeon $s$ had an above-median share of the cases at hospital $h$ in the prior quarter. This latter parameterization allows for the possibility that the relationship between SHARE and surgeon influence may not be continuous. Columns 3 and 4 provide analogous results using our second proxy for influence—CITATIONS. We note that the slight decline in the sample size for these regressions is due to the fact that some hospitals did not have any surgeons with academic citations, thereby leaving the CITATIONS variable unidentified.

In all regressions in Table 6, the coefficient on the influence proxy has the predicted negative sign, although it is insignificant at conventional levels. For the purpose of this study, the most critical result in Table 6 is that the relationship between the HOSPCASE and OTHCASE coefficients remains similar to that in our previous models. Specifically, the coefficients on HOSPCASE and OTHCASE are similar in magnitude and significance to those in Tables 4 and 5, and the difference between the coefficients is significant at the 10% level or better. Overall, these results suggest that, even after controlling for a surgeon's influence within a particular hospital, his or her volume *at that facility* still plays a significant role in determining performance while volume *at other facilities* continues to have only an insignificant effect. This finding points to the importance of familiarity—above and beyond influence—in explaining the hospital-specific performance of surgeons. The use of additional proxies to disentangle these effects more conclusively represents an avenue for future research.

## 6. Conclusion

The empirical setting considered in this paper is particularly well suited for examining firm specificity in the performance of freelancers, as it enables us to observe many individuals, each of whom is working within multiple organizations roughly simultaneously. Further, the well-established relationship between surgeon volume and clinical outcomes serves as a convenient means of correlating an individual's degree of contact with a given firm to his or her performance within that organization. We find a substantial degree of firm specificity in surgeon performance. More precisely, higher volume in a prior period for a given surgeon at a particular hospital is correlated with significantly lower risk-adjusted mortality for that surgeon-hospital pair. That volume, however, does not significantly improve the surgeon's performance at *other* hospitals, thus suggesting that surgeon performance is not fully portable across organizations.

Further, we find that firm specificity in surgeon performance is not simply an artifact of a surgeon's influence or power relative to other cardiac surgeons at a particular hospital. That is, this effect is not solely explained by the fact that a surgeon with high volume at a given hospital may be able to command

**Table 6 Logistic Regression with Volume Measures and Proxies for Surgeon Influence**

| | Dependent variable: Did CABG patient die in hospital? | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $HOSPCASE_{s,h,q-1}$: Surgeon cases at hospital (prior quarter) | −0.0085*** (0.0026) | −0.0101*** (0.0026) | −0.0105*** (0.0024) | −0.0104*** (0.0024) |
| $OTHCASE_{s,h,q-1}$: Surgeon cases at other hospitals (prior quarter) | −0.0019 (0.0037) | −0.0013 (0.0038) | −0.0017 (0.0037) | −0.0019 (0.0037) |
| $SURGRAMR_{s,q-1}$: Surgeon RAMR (prior quarter) | 0.8382* (0.4851) | 0.8619* (0.4867) | 0.9237* (0.4807) | 0.9188* (0.4798) |
| $HOSPRAMR_{h,q-1}$: Hospital RAMR (prior quarter) | −0.5570 (1.7725) | −0.4623 (1.7793) | −0.7377 (1.7753) | −0.7458 (1.7742) |
| $SHARE_{h,q-1}$: Surgeon share of hospital's cases (prior quarter) | −0.0044 (0.0028) | | | |
| $HIGH\_SHARE_{h,q-1}$: Indicator for above-median share of hospital's cases (within hospital) | | −0.0609 (0.0858) | | |
| $CITATIONS_{s,h,q}$: Surgeon share of academic CABG citations at hospital (includes all surgeons practicing at hospital in current quarter) | | | −0.0001 (0.0012) | |
| $HIGH\_CITATIONS_{s,h,q}$: Indicator for above-median citations (within hospital) | | | | −0.0417 (0.0815) |
| Constant | −3.7191*** (1.2843) | −3.7714*** (1.2839) | −3.6314*** (1.3003) | −3.6147*** (1.2994) |
| Level at which HOSPCASE is significantly different from OTHCASE (%) | 9 | 2 | 2 | 2 |
| Observations | 33,610 | 33,610 | 30,032 | 30,032 |
| Pseudo $R^2$ | 0.1766 | 0.1762 | 0.1762 | 0.1762 |
| Wald chi-squared | 2,053.4*** | 1,976.2*** | 1,830.0*** | 1,828.8*** |

*Notes.* *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The following variables are included in the regressions but are not shown in the table: age, $age^2/100$, fixed effects for quarter year, and indicators for cardiogenic shock, concurrent angioplasty, complicated hypertension, dialysis, female gender, heart failure, and prior CABG or heart valve surgery. Standard errors are heteroskedasticity robust and clustered by surgeon.

superior resources relative to her colleagues. Rather, this relationship also reflects the productive benefits associated with a surgeon's familiarity with critical assets of the hospital organization. The specific nature of these key organizational assets—which may be specific employees, team structures, or operating routines—represents an area for further research.

One question raised by our findings is why surgeons continue to split their time across hospitals if such behavior has a detrimental impact on performance. While the focus of this study is on the effects of splitting behavior rather than its causes, we can speculate as to some of the reasons for this activity. Above all, it is not clear how obvious these performance differences are to surgeons. First, a surgeon who does a large number of procedures in a given period may not be able to distinguish a 1% mortality rate from a 3% rate. Further, he or she will not have a sense of how those outcomes compare to those for other surgeons on a risk-adjusted basis. To obtain such risk-adjusted comparisons, a surgeon must rely on the public report cards, which are released with a lag of several years. In addition, surgeons do not receive differential compensation for a case depending on whether the patient survives. We do not mean to suggest that surgeons are driven solely by financial incentives. Rather, compensation is one way in which signals can be sent to surgeons about their performance and, in this setting, that signal is muted. In reality, surgeons do care about their outcomes, but it is not easy for them to determine whether a spike in their mortality rate is due to something particular to their performance or to a larger trend across all surgeons at the hospital or within the market. Beyond the inability of surgeons to recognize performance issues, there may be multiple reasons why patients or surgeons derive benefit from splitting. For example, patients may gain the convenience of receiving treatment close to their home while surgeons may gain access to a broader population of patients. In the final calculus, the cost associated with lower performance may not offset the value of these other benefits.

Below we discuss some potential limitations and extensions of our analysis. First, our results are based on only one type of freelance worker. It is, therefore, possible that the pattern of organization-specific performance that we identify for cardiac surgeons may not be as strong in other settings. For example, in cases where firms are exploring knowledge that is "technologically distant" (Song et al. 2003) from their current areas of expertise, there may be more benefits associated with splitting as a means of knowledge transfer. Offsetting this possible limitation of our analysis, however, is the benefit of the rich detail that our data provide about the degree of splitting activity for individual workers and the objective quality of their performance within different organizations.

Second, our attempts to disentangle the roles of familiarity and influence in explaining firm-specific performance would benefit from additional proxies for surgeon influence. Such measures might include hospital-specific surgeon tenure or additional information about the degree to which specific surgeons are active in other influential activities (e.g., board or committee memberships) either within the hospital or within wider professional associations.

Finally, our analysis might benefit from a deeper panel with additional years of data for each surgeon and hospital. This would allow us to investigate whether the degree of firm specificity changed as CABG technology matured over time. Pennsylvania did not publicly report information for the years from 1996 to 1999, thus creating gaps in efforts to create a continuous panel from 1994 to 2000. Even if such a panel were constructed, we would still not be able to examine the impact of a surgeon's cumulative volume—as opposed to our measure of recent volume—without having information dating back to a surgeon's first CABG procedure. Cumulative volume would enable us to make definitive statements about the nature of the volume-outcome effects that we observe for CABG. Specifically, it would help us separate volume-outcome effects due to increased familiarity over time from similar effects due to an idiosyncratically good match between a surgeon and a hospital. While pure matching could theoretically explain our result, practical examples of pure matching are rare in this setting. The more common anecdote is that a surgeon builds firm-specific ties over time rather than having them from day one. In addition, Benkard's (2000) findings on organizational forgetting in aircraft production suggest that "recent production is more important than more-distant past production in determining a firm's current efficiency" (p. 1037). Longitudinal information would allow us to examine not only the effect of volume increases on firm-specific improvement, but also the potential impact of volume reductions on firm-specific "forgetting." Regardless of whether our results are due to matching or increased familiarity over time, they still highlight the key point that the performance of pivotal freelancers does depend on organizational context.

Despite these limitations, this study has implications for managers both within and beyond the health care industry. First, it sheds light on potential limits to transferring knowledge or expertise across firms.[18] An individual firm can benefit from the knowledge residing in other firms either through informal discussion or interaction with members of that firm (Ingram and Simons 2002) or by acquiring the key human or physical assets in which this knowledge resides (Almeida and Kogut 1999, Rosenkopf and Almeida 2003, Song et al. 2003). Our analysis contributes to this latter group of studies, although our focus is on the transfer of performance rather than knowledge. Its findings should encourage managers to take a more critical eye to the practice of building firm capabilities through the "best-athlete" strategy of hiring. Particularly when a highly skilled worker must interact with a complex array of other assets—human and physical—within a given firm, the performance of that worker may not be easily transferred across organizational settings.

An online supplement to this paper is available on the *Management Science* website at http://mansci.pubs.informs.org/ecompanion.html.

**Acknowledgments**
The authors thank Gautam Ahuja, Diane Bailey, Michael Tushman, Sidney Winter, anonymous referees, the associate editor, and seminar participants at the Academy of Management, Boston University, Harvard University, Northwestern University, and the Wharton Technology Miniconference for helpful comments. The authors also thank Eli Strick for his excellent research assistance and acknowledge financial support from the Division of Research at Harvard Business School. The data used in this analysis were obtained from the Pennsylvania Health Care Cost Containment Council (PHC4), which issues the following disclaimer: The Pennsylvania Health Care Cost Containment Council (PHC4) is an independent state agency responsible for addressing the problem of escalating health costs, ensuring the quality of health care, and increasing access to health care for all citizens regardless of ability to pay. PHC4 has provided data to this entity in an effort to further PHC4's mission of educating the public and containing health care costs in Pennsylvania. PHC4, its agents and staff, have made no representation, guarantee, or warranty, express or implied, that the data—financial, patient, payor, and physician specific information—provided to this entity are error free, or that the use of the data will avoid differences of opinion or interpretation. This analysis was not prepared by PHC4. This analysis was done by Robert S. Huckman and Gary P. Pisano. PHC4, its agents and staff, bear no responsibility or liability for the results of the analysis, which are solely the opinion of the authors.

[18] See Tushman (1977), Darr et al. (1995), and Argote and Ingram (2000) for reviews of the literature on knowledge transfer between organizations.

## References


Adler, P., K. Clark. 1991. Behind the learning curve: A sketch of the learning process. *Management Sci.* **37**(3) 267–281.

Ahuja, G. 2003. When Atlas shrugged: Preemption, complexity and division of labor in a theory of appropriability. Working paper, University of Michigan, Ann Arbor, MI.

Allison, P., J. S. Long. 1990. Departmental effects on scientific productivity. *Amer. Sociol. Rev.* **55**(4) 469–478.

Almeida, P., B. Kogut. 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management Sci.* **45**(7) 905–917.

American Heart Association. 2001. *2002 Heart and Stroke Statistical Update*. American Heart Association, Dallas, TX.

Argote, L., D. Epple. 1990. Learning curves in manufacturing. *Science* **247**(4945) 920–924.

Argote, L., P. Ingram. 2000. Knowledge transfer: A basis for competitive advantage in firms. *Organ. Behavior Human Decision Processes* **82**(1) 150–169.

Argote, L., S. Beckman, D. Epple. 1990. The persistence and transfer of learning in industrial settings. *Management Sci.* **36**(2) 140–154.

Becker, G. 1962. Investment in human capital: A theoretical analysis. *J. Political Econom.* **70**(5) 9–49.

Benkard, C. L. 2000. Learning and forgetting: The dynamics of aircraft production. *Amer. Econom. Rev.* **90**(4) 1034–1054.

Berman, S., J. Down, C. Hill. 2002. Tacit knowledge as a source of competitive advantage in the National Basketball Association. *Acad. Management J.* **45**(1) 13–31.

Chillemi, O., B. Gui. 1997. Team human capital and worker mobility. *J. Labor Econom.* **15**(4) 567–585.

Cymrot, D., J. Dunlevy. 1987. Are free agent perspicacious peregrinators? *Rev. Econom. Statist.* **69**(1) 50–58.

Darr, E., L. Argote, D. Epple. 1995. The acquisition, transfer and depreciation of knowledge in service organizations: Productivity in franchises. *Management Sci.* **41**(11) 1750–1762.

Davis-Blake, A., B. Uzzi. 1993. Determinants of employment externalization: A study of temporary workers and independent contractors. *Admin. Sci. Quart.* **38**(2) 195–223.

Delaney, P., L. Reder, J. Staszewski, F. Ritter. 1998. The strategy-specific nature of improvement: The power law applies by strategy within task. *Psych. Sci.* **9**(1) 1–7.

Groysberg, B. 2001. Can they take it with them? The portability of star knowledge workers' performance: Myth or reality? Working paper, Harvard University, Cambridge, MA.

Hannan, E., J. O'Donnell, H. Kilburn, H. Bernard, A. Yazici. 1989. Investigation of the relationship between volume and mortality for surgical procedures performed in New York state hospitals. *J. Amer. Medical Assoc.* **262**(4) 503–510.

Hannan, E., H. Kilburn, H. Bernard, J. O'Donnell, G. Lukacik, E. Shields. 1991. Coronary artery bypass graft surgery: The relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. *Medical Care* **29**(11) 1094–1107.

Hayes, R., K. Clark. 1986. Why some factories are more productive than others. *Harvard Bus. Rev.* **64** 66–74.

Hayes, R., S. Wheelwright. 1984. *Restoring Our Competitive Edge: Competing Through Manufacturing*. John Wiley, New York.

Health Forum. 2002. *Hospital Statistics*. Health Forum, Chicago, IL.

Houseman, S. 2001. Why employers use flexible staffing arrangements: Evidence from an establishment survey. *Indust. Labor Relations Rev.* **55**(1) 149–170.

Huckman, R. 2003. The utilization of competing technologies within the firm: Evidence from cardiac procedures. *Management Sci.* **49**(5) 599–617.

Ichniowski, C., K. Shaw, G. Prennushi. 1997. The effects of human resource management practices on productivity: A study of steel finishing lines. *Amer. Econom. Rev.* **87**(3) 291–313.

Ingram, P., T. Simons. 2002. The transfer of experience in groups of organizations: Implications for performance and competition. *Management Sci.* **48**(12) 1517–1533.

Jarmon, R., A. Paulson, D. Rebne. 1998. Contractor performance: How good are contingent workers at the professional level? *IEEE Trans. Engrg. Management* **45**(1) 11–19.

Jovanovic, B. 1979. Job matching and the theory of turnover. *J. Political Econom.* **87**(5) 972–990.

Kahn, L. 1993. Free agency, long-term contracts and compensation in major league baseball: Estimates from panel data. *Rev. Econom. Statist.* **75**(1) 157–164.

Katz, R. 1982. The effects of group longevity on project communication and performance. *Admin. Sci. Quart.* **27** 81–104.

Krautmann, A., M. Oppenheimer. 1994. Free agency and the allocation of labor in major league baseball. *Managerial Decision Econom.* **15**(5) 459–469.

Long, J. S., R. McGinnis. 1981. Organizational context and scientific productivity. *Amer. Sociol. Rev.* **46**(4) 422–442.

Luft, H., J. P. Bunker, A. Enthoven. 1979. Should operations be regionalized? The empirical relation between surgical volume and mortality. *New England J. Medicine* **301**(25) 1364–1369.

MacDonald, D., M. Reynolds. 1994. Are baseball players paid their marginal products? *Managerial Decision Econom.* **15**(5) 443–457.

Mailath, G., A. Postlewaite. 1990. Workers versus firms: Bargaining over a firm's value. *Rev. Econom. Stud.* **57**(3) 369–380.

Matusik, S., C. Hill. 1998. The utilization of contingent work, knowledge creation, and competitive advantage. *Acad. Management Rev.* **23**(4) 680–697.

Mayer, K., J. Nickerson. 2005. Antecedents and performance implications of contracting for knowledge workers: Evidence from information technology services. *Organ. Sci.* **16**(3) 225–242.

Milgrom, P., J. Roberts. 1988. An economic approach to influence activities in organizations. *Amer. J. Sociol.* **94** S154–S179.

Milgrom, P., J. Roberts. 1990. The economics of modern manufacturing: Technology, strategy, and organization. *Amer. Econom. Rev.* **80**(3) 511–528.

Newell, A., P. Rosenbloom. 1981. Mechanisms of skill acquisition and the power law of practice. J. Anderson, ed. *Cognitive Skills and Their Acquisition*. Erlbaum, Hillsdale, NJ, 1–55.

Parsons, D. 1972. Specific human capital: An application to quit rates and layoff rates. *J. Political Econom.* **80**(6) 1120–1143.

Pennsylvania Health Care Cost Containment Council. 1998. *Coronary Artery Bypass Graft Surgery—1994–95 Data: Research Methods and Results*. Harrisburg, PA.

Pennsylvania Health Care Cost Containment Council. 2002. *Pennsylvania's Guide to Coronary Artery Bypass Graft Surgery, 2000*. Harrisburg, PA.

Peteraf, M. 1993. The cornerstones of competitive advantage: A resource-based view. *Strategic Management J.* **14** 179–191.

Pisano, G., R. Bohmer, A. Edmondson. 2001. Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Sci.* **47**(6) 752–768.

Podolny, J., T. Stuart. 1995. A role-based ecology of technological change. *Amer. J. Sociol.* **100**(5) 1224–1260.

Polanyi, M. 1966. *The Tacit Dimension*. Anchor Day Books, New York.

Reagans, R., L. Argote, D. Brooks. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Sci.* **51**(6) 869–881.

Rivkin, J. 2000. Imitation of complex strategies. *Management Sci.* **46**(6) 824–844.

Rosenkopf, L., P. Almeida. 2003. Overcoming local search through alliances and mobility. *Management Sci.* **49**(6) 751–766.

Showstack, J., K. Rosenfeldt, D. Garnick, H. Luft, R. Schaffarzick, J. Fowles. 1987. Association of volume with outcome of coronary artery bypass graft surgery: Scheduled vs. nonscheduled operations. *J. Amer. Medical Assoc.* **257**(6) 785–789.

Song, J., P. Almeida, G. Wu. 2003. Learning-by-hiring: When is mobility more likely to facilitate interfirm knowledge transfer? *Management Sci.* **49**(4) 351–365.

Storey, J., P. Quintas, P. Taylor, W. Fowle. 2002. Flexible employment contracts and their implications for product and process innovation. *Internat. J. Human Resource Management* **13**(1) 1–18.

Trajtenberg, M. 1990. A penny for your quotes: Patent citations and the value of innovations. *RAND J. Econom.* **21**(1) 172–187.

Tushman, M. 1977. Special boundary roles in the innovation process. *Admin. Sci. Quart.* **22**(4) 587–605.

Weick, K., K. Roberts. 1993. Collective mind in organizations: Heedful interrelating on flight decks. *Admin. Sci. Quart.* **38** 357–381.