Outcomes and Quality

Provider-Hospital “Fit” and Patient Outcomes: Evidence from Massachusetts Cardiac Surgeons, 2002–2004

Marco D. Huesch

Objective. To examine whether the “fit” of a surgeon with hospital resources impacts cardiac surgery outcomes, separately from hospital or surgeon effects.

Data Sources. Retrospective secondary data from the Massachusetts Department of Public Health’s Data Analysis Center, on all 12,983 adult isolated coronary artery bypass surgical admissions in state-regulated hospitals from 2002 through 2004. Clinically audited chart data was collected using Society of Thoracic Surgeons National Cardiac Surgery Database tools and cross-referenced with administrative discharge data in the Division of Health Care Finance and Policy. Mortality was followed up through 2007 via the state vital statistics registry.

Study Design. Analysis was at the patient level for those receiving isolated coronary artery bypass surgery (CABG). Sixteen outcomes included 30-day mortality, major morbidity, indicators of perioperative, and predischARGE processes of care. Hierarchical crossed mixed models were used to estimate fixed covariate and random effects at hospital, surgeon, and hospital $\times$ surgeon level.

Principal Findings. Hospital volume was associated with significantly reduced intraoperative durations and significantly increased probability of aspirin, β -blocker, and lipid-lowering discharge medication use. The proportion of outcome variability due to unobserved hospital $\times$ surgeon interaction effects was small but meaningful for intraoperative practices, discharge destination, and medication use. For readmissions and mortality within 30 days or 1 year, unobserved patient and hospital factors drove almost all variability in outcomes.

Conclusions. Among Massachusetts patients receiving isolated CABG, consistent evidence was found that the hospital $\times$ surgeon combination independently impacted patient outcomes, beyond hospital or surgeon effects. Such distinct local interactions between a surgeon and hospital resources may play an important part in moderating quality improvement efforts, although residual patient-level factors generally contributed the most to outcome variability.

Key Words. Quality of care, care processes, cardiac surgery, hierarchical modeling, mixed models

Do the processes of care that drive quality in a hospital interact with individual physicians? Such interactions might be separate from physician effects (e.g., intrinsic skill) or hospital effects (e.g., advanced technology or efficient processes). For example, a positive interaction effect could stem from a well-fitting surgical team where a cardiac surgeon is matched with distinct hospital-based resources (e.g., a particular operating room, familiar scrub nurses, or anesthesiologists).

If hospital–surgeon interactions did contribute meaningfully to the variability in patient outcomes, then improving quality would require more than monitoring and enhancing individual provider quality or fine-tuning hospital processes. An additional imperative would then be to match providers to hospital assets in teams that “fit” together better.

We hypothesize that the “fit” between hospital and surgeon could thus moderate the impact of the average process of care in a hospital or the impact of the average surgeon’s practice pattern. The potential sign of such moderation is not clear. For example, high familiarity among members of a surgical care team could improve outcomes (Pisano, Bohmer, and Edmondson 2001; Reagans, Argote, and Brooks 2005). Poor surgeon–hospital combinations might lead to stark deviations from best practice as in pediatric cardiac surgery at the Royal Infirmary (Kennedy Commission, 2001).

Scattered empirical evidence hints at such moderation. Recent work by Carey et al. (2008) suggests that a cardiac surgeon’s effect on mortality is less beneficial when working across multiple high-volume hospitals than when concentrating cases within one high-volume hospital. Other work finds that the outcomes of the same surgeon vary across different hospitals (Hannan et al. 2003; Huckman and Pisano 2006; Miyata et al. 2008). Case studies also point to the existence of interaction effects. For instance, the successful adoption of new minimally invasive techniques was moderated by surgeons’ styles of interacting with their operating teams (Edmondson et al. 2003).

We investigate these important questions using a novel approach and rich observational data on cardiac surgeries in Massachusetts. The patient records comprise clinically audited, chart data, allowing accurate and

Address correspondence to Marco D. Huesch, M.B.B.S., Ph.D., Duke University, The Fuqua School of Business, 100 Fuqua Drive, Box 90120, Durham, NC 27708-0120; e-mail: m.huesch@duke.edu. Dr. Huesch is with the Health Sector Management Area, Duke Fuqua School of Business. He is also with the Department of Community & Family Medicine, Duke School of Medicine, and the Department of Health Policy & Management, UNC Gillings School of Global Public Health.

comprehensive control for patient risk (Torchiana and Meyer 2005; Tsai et al. 2006).

Our use of hierarchical models extends prior use of random effects for physicians (Bronskill et al. 2002; Pietz et al. 2002; Ferraris et al. 2008), for hospitals (Hannan et al. 2003; Peterson et al. 2004; Carey et al. 2008; Marcin et al. 2008), or for both (Austin, Tu, and Alter 2003; Hirth et al. 2009). Our approach hinges on two-way crossed mixed models with random effects for hospital, surgeon, and the hospital $\times$ surgeon interaction. The interaction effects are identified because surgeons are not exclusively nested within hospitals.

Our outcomes of interest also extend beyond traditional endpoints of 30-day mortality, long-term follow-up mortality, and major morbidity events. We hypothesize that unobservable hospital $\times$ surgeon interactions could also impact perioperative measures such as total skin-to-skin, cross-clamp, and perfusion durations, and the use of internal mammary artery grafts, off-pump techniques, and particular discharge medications. These quality metrics are in keeping with the recent focus on processes of care in cardiac surgery (Shahian 2004; Shahian et al. 2007; Shahian and Normand 2008).

BACKGROUND

Our research question mirrors similar investigations in the social sciences. The economics and sociology disciplines contain extensive literatures on the “where and how” of organizational performance improvement. Learning using or doing are two accepted mechanisms of the “how.” As to the “where,” one perspective suggests that individual workers are the locus of shop-floor problem solving (Arrow 1962). A well-developed complementary perspective notes the importance of firm- or team-level effects (Becker 1962) and stresses the importance of distilling and codifying such improvements (Cohen and Levinthal 1990), especially given the “tacit” or “sticky” nature of such knowledge (von Hippel 1994). This interplay between knowledge, individuals, and firms can lead to a knowledge-based cospecialization of workers and the firm (Pennings and Wezel 2007).

Research Hypotheses

An analogous interplay exists between hospitals, their resources, individual cardiac surgeons, and the knowledge required to optimize patient outcomes. Much knowledge is codified in hospital procedures, routines, and practices and may be fungible across centers (O’Connor et al. 1996).

Hospitals influence the outcomes of patients through the fidelity of their care delivery processes, the quality of their human resources, and their technological sophistication. While this may be correlated with hospital scale or experience, observed hospital procedural volume may not capture all the influence of the hospital. Accordingly we hypothesize that:

H₁: Hospital-level factors contribute a portion of unexplained variance in patient outcomes, even after controlling for hospital procedural volume.

Surgeons are individual knowledge workers who differ in their possession of skills and clinical judgment in ways that might affect outcomes. Leading cardiac surgeons state that “there are differences in surgical skill among surgeons is something that we all know . . . but never discuss publicly” (Bonchek 2002, p. 433). Clinical judgment is indispensable, because consensus guidelines leave substantial room for individual judgment (Schofield 2003).

Delphi-style experiments using exemplar cases find poor agreement between cardiac surgeons (Kee et al. 1998; Denvir et al. 2006). Heterogeneous decision making may drive differences in cardiac revascularization rate across geographic areas, controlling for patient risk (Wennberg 1999). We thus expect:

H₂: Surgeon-level factors contribute a portion of unexplained variance in patient outcomes, even after controlling for surgeon procedural volume.

An interaction effect between surgeons and hospitals may also influence patient outcomes. Edmondson et al. (2003) document the performance benefits of team leadership style in cardiac surgery. The idiosyncratic formation of teams with well-fitting members in an operating team may endow some teams with greater joint working experience (Reagans, Argote, and Brooks 2005).

Interaction effects may also arise from the specialization of surgeon experience to the different hospitals they work at, due to familiarity with hospital technological assets (Huckman and Pisano 2006) or a surgeon’s idiosyncratic ability to gain an appropriate overview of patient inputs and outcomes within a hospital (Young et al. 1998). Hospitals are complex entities in which competing constituencies struggle for control of hospital resources and patients, especially in the treatment of ischemic heart disease where the turf war between cardiologists and surgeons is well known (Huckman 2003). These turf wars may impact the optimal therapeutic decision (Hemingway et al. 2001). It is plausible that some surgeons are better able to navigate

through such potential conflicts in some hospitals than others in other centers. On balance, therefore, our central hypothesis is:

H₃: Hospital × surgeon interaction factors contribute a portion of unexplained variance in patient outcomes, even after controlling for both hospital and surgeon procedural volumes.

ANALYTICAL STRATEGY

We obtained authorization under M.G.L. c.111, Section 24A to use Massachusetts Data Analysis Center (Mass-DAC) data from the Massachusetts’ Department of Public Health (DPH). These data are administered by Harvard Medical School and comprise all cardiac surgery admissions for patients aged 18 years and older in all 14 Massachusetts state-regulated hospitals over the period January 1, 2002 through December 31, 2004. A total of 22,435 records of cardiac surgery admissions by one of 61 surgeons were coded (usually by clinical staff) at the hospital using the Society of Thoracic Surgeons National Adult Cardiac Database tools. Data auditing and validation had been conducted by Mass-DAC, the DPH, and specialist cardiac surgeons.

To ensure that hospital reported 30-day mortality data were accurate and to allow follow-up of longer-term mortality out to December 31, 2007, Mass-DAC linked patient identifiers with death certificate data held by the Massachusetts Registry of Vital Records and Statistics. There was no follow-up for patients who moved to a different state or who left the country.

Over the 3 years of our data, there were three versions of software used to code data fields (Adult Cardiac software version 2.35, 2.41, and 2.52); coding and definition changes are documented in the Society of Thoracic Surgeons’ coding manuals. We used the DPH field (proctype = 1) to identify isolated coronary artery bypass surgical (iCABG) procedures: coronary artery bypass grafts performed without valve or other cardiac surgical procedure. We discarded the 9,452 records of patients not receiving an iCABG, and of the three surgeons who did not perform any iCABG procedures. Summary statistics were calculated from the remaining set of all 12,983 iCABG procedures performed by 58 surgeons. Missing values, inapplicable key dependent variables, and/or coding changes reduced the number of observations further in some regressions (Appendix Table SA3).

We used 64-bit *STATA/SE* v10.1 for all statistical analyses (all output logged in the Appendix Table SA4). We report only two-tailed tests of

significance, and considered p -values of $<.05$ as significant. This study was approved by the Duke University Health System Institutional Review Board.

Econometric Specification

We specified pooled two-way error components models with crossed random effects and an interaction term. Clustering is represented by random effects at within-surgeon ζ_{1s} , within-hospital ζ_{2h} , and within-hospital $\times$ surgeon dyad $\zeta_{3h \times s}$ levels. Patients are indexed i , surgeons s , and hospitals h . For continuous outcomes y , we estimate the following using restricted maximum likelihood (ML) or an iterated EM approximation to restricted ML, specifying $\varepsilon_{ish} \sim N(0, \theta)$:

$$y_{ish} = \alpha + \beta X_i + \gamma X_{h,-s} + \delta X_{s,h} + \varphi X_{s,-h} + \mu_t + \zeta_{1s} + \zeta_{2h} + \zeta_{3h \times s} + \varepsilon_{ish}$$

For binary outcomes, we estimated latent variable equations using a Laplacian approximation to ML and specifying the logistic density for the errors $\varepsilon_{ish} \sim \Lambda(0, \pi^2/3)$:

$$\begin{aligned} y_{ish}^* &= \alpha + \beta X_i + \gamma X_{h,-s} + \delta X_{s,h} + \varphi X_{s,-h} + \mu_t + \zeta_{1s} + \zeta_{2h} + \zeta_{3h \times s} + \varepsilon_{ish} \\ y_{ish}^* > 0 &\Leftrightarrow y_{ish} = 1 \\ \Pr\{y_{ish} = 1 | X, \mu, \zeta\} &= \Pr\{y_{ish}^* > 0\} = \Lambda(\alpha + \beta X_i + \gamma X_{h,-s} \\ &\quad + \delta X_{s,h} + \varphi X_{s,-h} + \mu_t + \zeta_{1s} + \zeta_{2h} + \zeta_{3h \times s}) \end{aligned}$$

Following the hierarchical models literature (Rabe-Hesketh and Skrondal 2008), we distinguish on the right-hand side between the fixed part and the random part. In the fixed part, we represent patient-level control covariates X_b , indicators μ_t for calendar years, and procedural iCABG volume per year covariates. We distinguish between the volume that the focal surgeon performs in the focal hospital ($X_{s,h}$), the volume that the focal surgeon performs elsewhere ($X_{s,-h}$), and the volume that the hospital performs not including the focal surgeon's volume ($X_{h,-s}$).¹

In the random part we collect the random effects ζ and specify the following distributions $\zeta_{1s} \sim N(0, \psi_1)$, $\zeta_{2h} \sim N(0, \psi_2)$, $\zeta_{3h \times s} \sim N(0, \psi_3)$ each conditionally independent of each other and covariates, and where θ and ψ are variances to be estimated. Note that the ζ in the continuous outcome equations are different to those in the binary outcome equations. Our key interest is the magnitude of the estimated variance ψ_s of the hospital $\times$ surgeon level random intercept and whether it accounts for a large proportion of the total unexplained variance (the sum $\psi_1 + \psi_2 + \psi_3 + \theta$, by the independence assumption). To provide graphical illustrations of the random effects, we

recover empirical Bayesian predictions of the mean or mode of the posterior density (Rabe-Hesketh and Skrondal 2008).² They are “shrunk” toward the zero mean of the prior density for smaller clusters (e.g., surgeons with fewer repeated observations), for noisy data (i.e., larger residual unexplained variance), and for informative priors (e.g., surgeons whose random intercept is estimated with smaller variance).

Dependent Variables

We study 16 practices and outcomes, but we were most interested in markers of intraoperative process efficiency and effectiveness. We coded for the gold standard binary events of internal mammary artery use (arterial graft material which lasts longer than vein grafts) and full off-pump operation (where possible, conducting the entire operation with a beating heart and without cardio pulmonary bypass). The broad consensus within the specialty is that off-pump techniques reduce intraoperative damage to the great vessels (e.g., dislodging plaque from the ascending aorta) and reduce the risk of damage to the patient’s blood (e.g., clotting or broken red blood cells causing subsequent kidney damage).³ We also calculated continuous variables representing the duration of total incision,⁴ perfusion, aortic cross-clamp, and ventilation.

We also coded postoperative major morbidity for any one of the following complications: deep sternal wound infection, postoperative renal failure, postoperative stroke not including coma, prolonged mechanical ventilation > 24 hours, and postoperative reinterventions (for bleed, tamponade, valvular dysfunction, graft occlusion, or other reasons). Separately we coded postoperative myocardial infarction (MI).

Discharge medication use was coded for aspirin, lipid lowering therapy, and β -blocker prescription. We also considered the following traditional outcomes: mortality within 30 days postoperatively, readmission to the same hospital for any reason within 30 days of operation, and mortality within 12 months of surgery.⁵ As our long-term mortality data do not include patients who moved and died out of state or out of country, inferences regarding this outcome are necessarily limited to Massachusetts residents.

In two specification checks, we tested to ensure that the model was not producing artifactual results. We picked the arbitrary “outcome” of surgery taking place on a Tuesday for nonelective patients because elective patients tend not to be scheduled for weekends and Monday exhibits pent-up weekend admission demand. We also chose patient weight as an arbitrary outcome. We expected these models would poorly fit the data, that the key volume

covariates would not be significant, and that unsystematic variation attributable to patients would dominate the variance components.

Control Covariates

We use 31 clinical covariates previously shown to control for patient preexisting risk in mortality models (Peterson et al. 2004). We control for *demographics* (age, female gender, and nonwhite race), *family history* (coronary artery disease), and *comorbidities* (past/current smoking status, diabetes, hypercholesterolemia, renal failure with or without dialysis, hypertension, cerebrovascular accident and disease, infectious endocarditis, chronic lung disease, immunosuppression, and peripheral vascular disease).

Biometric and biomechanical factors were included (weight, height, creatinine, and left ventricular ejection fraction), and we coded for *prior cardiovascular interventions* (prior valve, CABG, or percutaneous coronary intervention within 6 hours of surgery). We also controlled for *cardiovascular preoperative comorbidities* (MI within 6 hours/1 day/7 days/> 7 days before surgery, congestive heart failure, unstable angina, cardiogenic shock, arrhythmia comprising ventricular tachycardia or ventricular fibrillation, heart block, atrial fibrillation, or atrial flutter) (Table 1).

Finally our data contained *indicators of disease severity* (triple vessel disease, left main disease > 50 percent, New York Heart Association Class II/III/IV disease) and *preoperative indicators of acuity and severity* (preoperative status of urgent/emergent/salvage operation, inotropic therapy, resuscitation required, intraaortic balloon pump in situ preoperatively).

To check the discriminatory power of our model we ran a diagnostic probit model of these covariates and dummy variables for year of surgery on 30-day mortality. We found a reasonable pseudo- R^2 of 16.6 percent and a C-index of 0.82.⁶ We document the model's satisfactory performance in specific higher risk groups with pseudo- R^2 never below 15 percent and C-index never below 0.81 (see Appendix Table SA5).

Summary Statistics

The demographic and clinical features of the patient population are reported in Table 1. The number of Massachusetts surgeons and hospitals was stable (Appendix Tables SA1 and SA2). Of the 14 hospitals performing cardiac surgery, the mean 3-year hospital caseload of iCABG procedures was 927 (range 145–1,860). The modal number of surgeons performing iCABG in

Table 1: Baseline Characteristics of the Study Population

<i>Variable</i>	<i>Mean</i>	<i>SD</i>
Demographics		
Age (years)	66.7	10.7
Female gender	25.8	43.8
Non-white race	9.0	28.7
Acuity and severity		
Urgent status	64.7	47.8
Emergent status	3.2	17.7
MI before surgery: 0–6 hours	0.9	9.4
7–24 hours	2.3	15.0
PCI before surgery: 0–6 hours	1.1	10.4
Preoperative inotropic support	1.6	12.5
Preoperative resuscitation	0.4	6.6
Congestive heart failure	17.9	38.4
Cardiogenic shock	1.7	12.8
Preoperative IABP	10.8	31.1
Triple vessel disease	77.9	41.5
Left main disease	33.8	47.3
Renal failure and dialysis	1.4	11.6
Renal failure no dialysis	5.4	22.5
NYHA Class IV	25.8	43.8
Biometric/biomechanical		
Weight (kg)	84.2	17.6
Height (cm)	170.8	9.9
Last creatinine (level)	1.1	0.8
Left ventricular ejection (fraction)*	49.1	13.6
Cardiovascular history		
Prior CABG	3.2	17.6
Prior valve operation	0.3	5.1
Peripheral vascular disease	17.7	38.2
Cerebrovascular disease	13.8	34.5
Cerebrovascular accident	7.1	25.7
Unstable angina	45.9	49.8
Hypertension	79.6	40.3
Hypercholesterolemia	80.7	39.5
Arrhythmia: VT/VF	2.1	14.2
Atrial fibrillation/flutter	6.3	24.3
Current smoker	16.5	37.1
Family history of CAD	37.0	48.3
Diabetes	37.7	48.5
Chronic lung disease	12.9	33.5
Immunosuppression status	2.6	15.8
Calendar time		
2003	33.8	47.3
2004	30.7	46.1
Operative techniques		
Full off-pump without conversion	11.1	31.4
Arterial grafts (number)	1.2	0.6
Venous grafts (number)	2.1	1.0
Single IMA used	91.6	27.8
Double IMA used	1.9	13.6

continued

Table 1. *Continued*

<i>Variable</i>	<i>Mean</i>	<i>SD</i>
Blood products used	56.4	49.6
Reintubated*	4.1	19.9
Durations		
Skin-to-skin time (minutes)*	270.4	81.6
Perfusion time (minutes)*	107.4	36.8
Cross clamp time (minutes)*	78.2	29.1
Ventilation (hours)*	29.0	125.5
Complications		
Major morbidity†	15.8	36.4
Perioperative MI	3.8	19.1
Postoperative pneumonia	3.5	18.3
Postoperative arrest	1.5	12.3
Stay and destination		
Length of stay‡	10.2	9.1
Discharged: directly home	63.8	48.1
Discharge to extended care	27.8	44.8
Discharged to nursing home	1.6	12.5
Discharge medications		
Aspirin§	87.2	33.4
ACE inhibitors§	21.1	40.8
β-Blockers§	82.7	37.8
Lipid lowering therapy§	72.0	44.9
Short-term outcomes		
Discharge mortality	1.9	13.8
Readmission <30 days§	10.8	31.1
Mortality <30 days	2.2	14.5
Long-term follow-up		
Mortality <12 months¶	5.7	23.2

Note. Patient-weighted statistics in percentages unless otherwise stated, for the 12,983 isolated CABG (iCABG) cases in MA in 2002–2004, except for the following:

*Differences due to Society of Thoracic Surgeons National Cardiac Database software changes, missing data, and applicability of measure. Affects Ejection Fraction (12,416), Reintubation (12,206), Total Incision Time (12,136), Perfusion Time (11,340), Cross-ClampTime (10,813), and Ventilation Hours (12,974);

†Defined as reintervention, stroke, renal failure, ventilation >24 hours, deep sternal infection;

‡Only for 8,997 patients due to Massachusetts Department of Health masking of dates at month level in 2004; and

§Calculated for 12,731 patients discharged alive due to the competing risk of death;

¶Defined as 12 full months survived after month of surgery not 365 days following day of surgery due to data masked at month level.

ACE, angiotensin converting enzyme; CAD, coronary artery disease; IABP, intra-aortic balloon pump; iCABG, isolated coronary artery bypass graft surgery; IMA, internal mammary artery; MI, myocardial infarction; PCI, percutaneous coronary intervention; VT/VF, ventricular tachycardia/fibrillation.

each hospital was 4 (range 2–9). The mean number of hospitals in which each surgeon performed iCABG was 1.3 (range 1–3).

Of the 56 surgeons included in our analysis (with total panel iCABG cases > 20), the mean total iCABG cases performed was 231.7 (range 23–471), while the mean other adult cardiac cases was 168.7 (6–802), for a mean total of 400.4 adult cardiac cases (range 41–906). One in four of these 56 surgeons worked across more than one hospital, collectively performing 27.6 percent of all iCABG, and on average seeing 255.4 iCABG cases (43–437). The remaining 42 surgeons who restricted their activities to only one hospital averaged 223.8 iCABG cases (23–471).

FINDINGS

Our substantive analysis uses estimated variances of random effects, not the actual random effects themselves. To illustrate the latter, we plot these in Figures 1 and 2 using the major morbidity models as examples. The random effects are predicted postestimation and correspond to the “average” patient. In our data the typical patient was a white male of 67 years with triple vessel disease presenting in urgent status.

In Figure 1 each marker represents one of the 72 distinct hospital–surgeon pairs. Horizontal lines represent common hospital effects, and vertical lines represent common surgeon effects. Figure 2 shows the effect of a particular hospital $\times$ surgeon combination over and above what is contributed by the hospital and the surgeon component. The surgeon and hospital $\times$ surgeon effects lying on the same approximately 45° line represent the surgeons who do not work across more than one hospital, and where a hospital $\times$ surgeon interaction effect cannot be identified separately from a surgeon effect.

For this binary outcome, the units are in additions to the logit of major morbidity. The patient-weighted rate of major morbidity is 15.8 percent, corresponding to odds of 0.158/(1–0.158) or a little < 1:5. A random effect of, say, +0.2 increases the logit by 0.2, all else constant. Exponentiation of 0.2 yields an increase in the odds by a factor of 1.22; the new odds are tighter, at a little < 1:4. Dividing these odds by one plus the odds yields the new higher major morbidity probability of 18.6 percent, nearly 3 percent points higher.

Procedural Volume

Table 2 reports estimates for the procedural volume covariates (Appendix Table SA4 has estimates of control covariates). Since our study spans 3 years

Figure 1: Surgeon and Hospital Random Effects for Major Morbidity Logistic Mixed Model

only, we are unable to distinguish between such procedural volume as representing scale or experience (Gaynor, Seider, and Vogt 2005). A linear probability model regression on 30-day mortality with just the fixed part of the model failed to demonstrate substantial multicollinearity among the fixed covariates (Appendix Table SA5).

There was no meaningful procedural volume association for traditional measures of readmission or mortality regardless of 30-day or 1-year threshold. A one case increase in hospital volume was associated with an increase in probability of 30-day mortality (odds ratio 1.002, $p < .05$). For outcomes in which learning by doing or repetition intensity might be important (e.g., total incision time), we found a small, precisely estimated effect of the expected sign for hospital caseload (all $p < .01$). Surgeon procedural volume was not a significant predictor of these measures of efficiency.

We found a strongly significant increase in the odds of appropriate discharge medication with hospital volume. We note that these are outcomes in

Figure 2: Surgeon and Hospital $\times$ Surgeon Interaction Random Effects for Major Morbidity

Units: e.g. 0.5 adds 0.5 to the logit of the outcome compared to the mean surgeon random effect of 0.

Note: Collinearity line is not aligned at 45 degrees due to biases in estimating variance components.

which hospital policies, pharmacy processes, and computerized physician order entry (CPOE) systems could play a large part. Surgeon procedural volume was associated with small but significant decreases in the likelihood of patients’ receiving optimal discharge therapy. Whether these could be congestion and attention issues or variation in physician compliance with CPOE (Isaac et al. 2009) is unclear.

Random-Effects Variance and Intraclass Correlations

In Table 2 we also report the estimated variance of the random effects. The first rows show the estimated variances ψ_1 , ψ_2 , and ψ_3 corresponding to the random effects ζ_{1s} , ζ_{2hs} , $\zeta_{3h \times s}$ and the residual patient-level variance θ corresponding to the residual ε_{ish} .⁷ By assumption of conditional independence, the variances are additively separable. This allows calculation of intraclass correlation (ICC).

Table 2: Hierarchical Models: Contribution of Hospital, Surgeon, and Hospital $\times$ Surgeon Interaction Effects to Outcome Variance

Operative Techniques			Durations			Complications	
IMA Used (1/0)	Full Off- Pump (1/0; LPM)	Perfusion (Minutes)	Cross-Clamp (Minutes)	Skin-Skin (Minutes)	Ventilation (Log Hours)	Major Morbidity (1/0)	Perioperative MI (1/0)
Fixed part							
Controls							
Procedural volume (iCABG cases/year)							
Hospital	1.000	-0.035	-0.021	-0.057	0.002	1.002	0.995
<i>t</i> -test	(+0.27)	(-4.20)***	(-3.02)**	(-3.47)***	(+8.95)**	(+2.81)**	(-3.77)***
Surgeon, focal hospital							
<i>t</i> -test	0.999	-0.011	0.007	-0.019	0.001	0.999	0.997
	(-0.29)	(-0.57)	(+0.45)	(-0.51)	(+1.73)*	(-0.39)	(-0.94)
Surgeon, elsewhere							
<i>t</i> -test	1.002	0.073	0.035	0.031	0.000	1.000	1.009
	(+0.53)	(+1.99)*	(+1.29)	(+0.44)	(-0.21)	(+0.07)	(+1.49)
Random part							
Random effects variances							
Hospital	0.11	113	77	362	0.16	0.12	5.33
Surgeon	0.22	282	262	884	0.01	0.08	0.38
Hospital × surgeon	0.00	25	0	144	0.01	0.05	0.00
Patient (residual)	3.29	789	477	3,094	0.52	3.29	3.29

Intraclass correlations (%)										
Hospital	3.1	5.3	9.3	9.5	8.1	23.0	3.3	59.3		
Surgeon	6.2	6.8	23.4	32.1	19.7	1.0	2.1	4.2		
Hospital × surgeon	0.0	21.7	2.0	0.0	3.2	2.0	1.5	0.0		
Patient (residual)	90.7	66.2	65.3	58.4	69.0	74.0	93.0	36.6		
Model fit										
Log (restricted) likelihood	−2,408	−1,076	−51,393	−46,493	−62,663	−13,398	−4,667	−1,493		
Wald test	935***	521***	1,431***	1,352***	1,491***	2,712***	1,083***	363***		
LR hierarchical model test	110***	4,023***	3,448***	4,066***	3,434***	1,191***	129***	322***		
Observations	12,398	12,398	10,795	10,306	11,517	12,060	12,398	12,398		

Notes. Estimated coefficients and tests significant at:

*** $p < .001$, ** $p < .01$, * $p < .05$, $^{\dagger}p < .10$ levels.

Perfusion, cross-clamping restricted to <300 minutes, skin-to-skin <600 minutes, ventilation <300 hours. See text for differences in number of observations. Linear models estimated using restricted maximum likelihood (REML). Logistic models estimated by Laplacian approximation to maximum likelihood except “Full off-pump” model estimated as linear probability model using REML.

iCABG/year, isolated annual coronary artery bypass graft cases in year of focal patient’s surgery. Fixed part of all models includes 31 (see Appendix Table SA3) risk factors, calendar year dummies, and a constant. IMA, internal mammary artery; MI, myocardial infarction.

Table 2. Continued

	Stay and Destination		Discharge Medications			30-Day Outcomes		1-Year Outcome	
	Stay Length (Log Days)	Direct Home (1/0)	Aspirin (1/0)	Lipid Lowering (1/0)	β-Blocker (1/0)	Readmit (1/0)	Mortality (1/0)	Mortality (1/0)	
Fixed part									
Controls									
Procedural volume (iCABG/year)					Included in all models				
Focal hospital	0.000	1.005	1.016	1.012	1.012	1.000	1.002	1.001	
t-test	(− 0.56)	(+6.73)***	(+12.69)***	(+14.45)***	(+13.31)***	(0.14)	(+2.04)*	(+1.29)	
Surgeon, focal hospital	− 0.001	1.004	1.005	1.006	1.002	0.999	1.002	1.001	
t-test	(− 2.58)**	(+3.17)**	(+2.07)*	(+3.49)***	(+0.93)	(− 1.14)	(+0.68)	(+0.46)	
Surgeon, elsewhere	0.000	0.998	1.002	1.001	0.995	1.000	0.994	0.998	
t-test	(− 0.94)	(− 0.88)	(+0.29)	(0.31)	(− 1.28)	(0.09)	(− 0.90)	(− 0.56)	
Random part									
Random effects variances									
Hospital	0.01	0.78	8.45	3.90	3.86	0.11	0.10	0.04	
Surgeon	0.01	0.00	0.00	0.21	0.25	0.00	0.00	0.00	
Hospital × surgeon	0.00	0.11	0.96	0.39	0.69	0.01	0.06	0.00	
Residual (patient)	0.16	3.29	3.29	3.29	3.29	3.29	3.29	3.29	
Intraclass correlations (%)									
Hospital	4.9	18.7	66.6	50.1	47.7	3.3	3.0	1.1	
Surgeon	5.3	0.0	0.0	2.7	3.1	0.0	0.0	0.0	
Hospital × surgeon	0.0	2.5	7.5	5.0	8.6	0.2	1.8	0.0	
Patient (residual)	89.8	78.8	25.9	42.2	40.6	96.5	95.2	98.9	
Model fit									
Log (restricted) likelihood	− 4.603	− 5.997	− 3.556	− 6.456	− 4.725	− 4.054	− 1.060	− 2.280	
Wald test	3,314***	2,396***	922***	1,154***	774***	174***	399***	778***	
LR hierarchical test	369***	232***	1,062***	893***	990***	81***	10*	5	
Observations	8,482	12,161	12,161	12,161	12,161	12,161	12,398	12,398	

Notes: Estimated coefficients and tests significant at:

*** $p < .001$, ** $p < .01$, and * $p < .05$ levels.

Fixed part of all models includes 31 (see Appendix Table SA3) risk factors, calendar year dummies, and a constant. Stay length model restricted to <51 days. Direct home, discharge medications, and readmit outcomes restricted to patients who survived discharge. See text for explanation of differences in number of observations. Logistic models estimated by Laplacian approximation to maximum likelihood. Linear models estimated using restricted maximum likelihood.

iCABG, isolated coronary artery bypass graft cases in year of focal patient's surgery; β , beta-blocker therapy.

The first two ICC of interest are $\rho_s = \psi_1/(\psi_1 + \psi_2 + \psi_3 + \theta)$, which estimates the correlation between outcomes of patients sharing the same operating surgeon but undergoing iCABG at different hospitals; and $\rho_h = \psi_2/(\psi_1 + \psi_2 + \psi_3 + \theta)$, which conversely estimates the outcome correlation between patients receiving iCABG at the same hospital but from different operating surgeons.

Hospital and Surgeon Effects

As we expected in H₁, hospitals make a small but meaningful contribution to patient variability in nearly all outcomes. This was especially clear for processes of care such as discharge medications. Here hospitals contributed between 48 and 68 percent of the variability captured by the random portion of the models, depending on therapy (Table 2).

For the perioperative MI complication, we estimated a large contribution (59 percent) of the hospital to the variability in this outcome even after controlling for higher hospital scale or longer experience (captured in the fixed part of the model as annual procedural volume).

We also found support for the second hypothesis on the impact of unobserved surgeon effects, especially for outcomes, which were most plausibly under control of the surgeon. Variability in use of the internal mammary and minimally invasive techniques was approximately 6 percent attributable to an unobserved surgeon effect. Intraoperative durations were characterized by a high surgeon effect (total variance contribution ranging from 20 to 32 percent), after controlling for scale or learning effects using surgeon procedural volume.

On the other hand, for outcomes such as extubation terminating ventilation time (affected by intensivists) the surgeon effect (0.8 percent) was far smaller in magnitude. For discharge medication variability, there was essentially no surgeon-level contribution. For traditional outcomes such as 30-day or 1-year mortality, there was little hospital and no surgeon-level effect seen.

Interaction Effects

The third hypothesis, H₃, on hospital $\times$ surgeon interactions was our central research question, and it was weakly supported. The interaction ICC, or $\rho_{h \times s} = \psi_3/(\psi_1 + \psi_2 + \psi_3 + \theta)$, estimates the average outcome correlation between patients with the same facility and surgeon *over and above* the correlation in outcomes due to such patients sharing the same surgeon or the same hospital.

Generally, we found the magnitude of the hospital $\times$ surgeon effect to be of far smaller magnitude to that of the hospital or surgeon effects themselves.

For example, in operative techniques and intraoperative durations, we found that the interaction effect drives between 0.0 and 3.2 percent of outcome variance, except for use of off-pump techniques. For this technically demanding minimally invasive operative approach, we found a very high interaction effect (22 percent contribution to total variance). We explicate this as being due to idiosyncratic hospital practices, technology, and experience interacting with surgeon propensities and abilities. As Pisano, Bohmer, and Edmondson (2001) report, minimally invasive technology appears highly susceptible to teamwork dynamics within the operating room.

Similar interaction effects are plausible for discharge medications, where the interaction effect contributed between 5 and 9 percent of variability after controlling for patient preexisting risk, disease severity, etc. and after accounting for hospital and surgeon procedural volumes. We speculate that different institutional-based interventions to prescribe such medications have idiosyncratic effects on some physicians (Isaac et al. 2009).

In Table 3 we turn to two falsification tests in which, as predicted, none of the procedural volume covariates were significant in either model, and almost all of the unexplained variability in the random part of the model was driven by unsystematic residual patient factors.

Two findings merit further explanation. We found no interaction effect for the complication of infarction in the period between operation and discharge; the variability was driven very largely by hospital effects with a small surgeon effect. Hospital processes of care (e.g., pharmaceutical care) may plausibly take place with little direct hospital–surgeon interaction.

On the other hand, we noted that for the traditional 30-day mortality or readmission outcomes, there was essentially no individual surgeon effect, but our models did estimate a small interaction effect. We interpret this as supporting our central thesis. Even when surgeons do not differ greatly in their impact on outcomes (low variance of surgeon effects), their “fit” with the hospital, their fellow team members there, and the processes and technology there all still matters.

Finally, we note that the percentage contribution to overall variability of results was generally small for the sum of the estimated effects. Inspection of Table 2 shows that the residual, unexplained, patient-level variance had a low of 26 percent (aspirin discharge) and 37 percent (perioperative MI). More typical values were around 60 percent with a high of 97–99 percent (readmission within 30 days and 1-year mortality). Even for a well-understood mature, high-volume procedure such as CABG, and even with excellent data, much outcome variability appears unsystematic.

Table 3: Falsification Tests

	<i>Tuesday Surgery Nonelective Only (1/0)</i>	<i>Admit Weight All Patients (kg)</i>
Fixed part		
Controls	Included in both models	
Procedural volume (iCABG/year)		
Focal hospital	1.000	0.001
<i>t</i> -test	(− 0.02)	(+ 0.46)
Surgeon, focal hospital	1.000	0.002
<i>t</i> -test	(+ 0.09)	(+ 0.36)
Surgeon, elsewhere	1.003	− 0.009
<i>t</i> -test	(+ 1.13)	(− 0.74)
Random part		
Random effects variances		
Hospital	0.00	0.93
Surgeon	0.13	0.00
Hospital × surgeon	0.00	0.15
Residual (patient)	3.29	290
Intraclass correlations (%)		
Hospital	0.0	0.3
Surgeon	3.7	0.0
Hospital × surgeon	0.0	0.1
Patient (residual)	96.3	99.6
Model fit		
Log (restricted) likelihood	− 2,737	− 52,740
Wald test	54	900***
LR hierarchical model test	46***	18***
Observations	5,762	12,401

Notes. Estimated coefficients and tests significant at:

*** $p < .001$ level.

Day of the week only reported for 2002 and 2003, lowering observations. Fixed part of “Surgery on Tuesday” models includes 31 (see Appendix Table SA3) risk factors, calendar year dummies, and a constant. Estimated as logistic model by Laplacian approximation to maximum likelihood. Fixed part of “Weight” model excludes weight, height, and age covariates in control list. “Weight” model estimated as linear model using restricted maximum likelihood.

iCABG, isolated coronary artery bypass graft cases in year of focal patient’s surgery.

DISCUSSION

Our study of interaction effects among cardiac surgeons demonstrated that a small but meaningful contribution to important patient outcomes is associated with the hospital × surgeon match. This effect was separate from that of the surgeon himself or herself and from that of the hospital where he or she practiced. Hospital procedural volume was a small but highly significant predictor of many patient process outcomes. Lastly, we showed that variability is

driven mostly by observed and unobserved patient-level factors, especially for key mortality outcomes.

Limitations

This study had several major limitations arising from the modeling approach, estimation methods, and setting. In the random part of our mixed models, we assumed that the random effects are time invariant and conditionally independent from each other and all the model covariates. In the fixed part, we estimated a single pooled coefficient for covariates for computational feasibility. This allows no differences, for instance, across surgeons in a procedural volume covariate's marginal effect beyond ones already reflected in the intercept. If the model is misspecified in this way, then the estimates of the covariates' association with the dependent variable are susceptible to an omitted variable bias (Neuhaus and Kalbfleisch 1998; Berlin et al. 1999).

We restricted reported analysis to iCABG cases, a homogenous population with validated risk adjustments. However, this approach discards a large number of other cardiac surgery cases such as valve replacements and also implies that the fixed volume variable coefficients may understate or misstate the effects of experience. In limited robustness tests (see Appendix Table SA4), we reestimated discharge medication use models using all adult cardiac cases for procedural volumes and found results similar to those reported here.

Our findings were not robust to endogeneity concerns in the association between surgeon or hospital effects and patient outcomes (Luft, Hunt, and Maerki 1987) and our data prevented an instrumental variables strategy.⁸ However, it seems implausible that patients can observe the idiosyncratic differences between particular hospital $\times$ surgeon combinations and select for these.

Our empirical setting allowed us to use clinically audited and rigorously defined data, this was only available to us from one (small) state. How this affects external validity is unclear. Given ongoing, leading edge quality improvement initiatives in Massachusetts, it may be that these results reflect an upper bound on interaction effects in other settings. On the other hand, given the excellent patient outcomes already achieved in Massachusetts, the results here may reflect more of a lower bound for other centers.

Conclusions

We see several policy implications of these results in our empirical setting. The first is that we ought to expect a large degree of variability in patient outcomes

to remain even after successful efforts to enhance intrinsic surgeon and hospital abilities or the fit between the two sides.

The second implication is the converse of the above point. Our results support the current focus on processes of care as a fruitful path to improving health care quality (Shahian 2004; Hales and Pronovost 2006; Gawande 2007; Shahian et al. 2007). We show small, but meaningful variation in such process outcomes between and within all levels of the hospital/surgeon/hospital $\times$ surgeon/patient hierarchy. For perioperative MI and dispensing of appropriate discharge medications, high variability appeared due to hospital-level factors. For other outcomes, such as some operative techniques and durations, surgeon and hospital $\times$ surgeon interactions appeared to play an important role. Especially for the interaction effect between surgeons and hospitals, future work in the field should attempt to isolate the mechanisms by which such matches impact patient outcome quality.

The third implication is that a policy focus on minimum procedural volume is not supported by our results. Past recommendations on cardiac volume thresholds by payor advocates such as the Leapfrog group have been criticized for missing the excellent outcomes obtained by many small-scale centers (Shahian and Normand 2003). Others have commented that hospital volume may be confounded by surgeon volume (Birkmeyer et al. 2003; Peterson et al. 2004). We find little evidence that Massachusetts centers exhibit meaningful returns to scale except in discharge medication use. This may still be confounded with larger hospitals having introduced advanced CPOE technology, which is designed to reduce medication variability.

Finally, the last implication of our findings is the sensitivity of policy to mechanism by which the hospital $\times$ surgeon interactions arise. Our analytical strategy did not isolate such mechanisms which may be more from idiosyncratic matching or may be driven more by a familiarity stemming from increased surgeon–hospital volume. If the former, then it would be important to understand how such idiosyncratic matches could be made in a more optimal manner. Future research is needed to explore this difficult normative question in empirical settings in which matches are exogenously changed (e.g., during the rotations of cardio-thoracic residents and fellows through Harvard’s teaching hospitals, or following the closure of St. Vincent’s in New York).⁹

If, on the other hand, the mechanism is driven more by familiarity between surgeons and team members, administrators, or hospital technology, then it would be important to understand whether strengthening such familiarity through concentration of volume might help. A policy of, say, centralization of “subscale” hospitals could have different consequences depending

on which mechanism was the dominant one. Such policy could possibly create benefits if familiarity was strengthened, but it could possibly create harm if older, better idiosyncratic matches were disrupted and newer, worse matches were formed.

NOTES

1. We are grateful to a referee for suggesting this disaggregation.
2. For the linear (nonlinear) models *STATA* estimates are best linear unbiased estimates of the *mean (mode)* of the posterior density of ζ_{xxx} (Rabe-Hesketh and Skrondal, 2008).
3. We are aware of the current, unresolved debate on whether a typical cardiac surgeon with typical volumes is able to achieve better outcomes using off-pump techniques.
4. Total incision time was computed by us from incision start time and stop time fields. The latter field changed definitions between *STSA CD* software versions 2.42 (nearest 30 minutes) and 2.52 (nearest minutes). Most timings were to the nearest minute, regardless of software version.
5. For 2004, DPH masked the mortality date at the month level. Year survival was defined as surviving 12 full months after the month of surgery.
6. The Hosmer–Lemeshow test was significant. However, in datasets this large the test is nearly always significant, even with almost perfect fit (Kramer and Zimmerman, 2007).
7. In models with dichotomous dependent variables, θ is constrained to be $\pi^2/3$ or ~ 3.29 by our construction of a latent variable logistic model.
8. Surgeon and hospital identifiers were masked, as were patient zipcodes, so we were unable to generate geographical or choice-model instruments to mitigate such omitted variable biases.
9. We are grateful to a referee for pointing out this important limitation of our study. Similar questions arise in the economics of matching in the marriage market, matching between firms and employers in general (Hao, 2008).

REFERENCES

- Arrow, K. J. 1962. “The Economic Implications of Learning by Doing.” *Review of Economic Studies* 29 (3): 155–73.
- Austin, P. C., J. V. Tu, and D. A. Alter. 2003. “Comparing Hierarchical Modeling with Traditional Logistic Regression Analysis among Patients Hospitalized with Acute Myocardial Infarction.” *American Heart Journal* 145: 27–35.
- Becker, G. S. 1962. “Investment in Human Capital.” *Journal of Political Economy* 70 (5): 9–49.
- Berlin, J. A., S. E. Kimmel, T. R. T. Have, and M. D. Sammel. 1999. “An Empirical Comparison of Several Clustered Data Approaches under Confounding Due to

- Cluster Effects in the Analysis of Complications of Coronary Angioplasty" *Biometrics* 55 (2): 470–6.
- Birkmeyer, J. D., T. A. Stukel, A. E. Siewers, P. P. Goodney, D. E. Wennberg, and F. L. Lucas. 2003. "Surgeon Volume and Operative Mortality in the United States." *New England Journal of Medicine* 349: 2117–2.
- Bonchek, L. I. 2002. "Off-Pump Coronary Bypass: Is It for Everyone?" *Journal of Thoracic and Cardiovascular Surgery* 124: 431–4.
- Kennedy Commission. 2001. Bristol Inquiry Final Report [accessed on October 20, 2009]. Available at http://www.bristol-inquiry.org.uk/final_report/report/Summary2.htm
- Bronskill, S. E., S.-L. T. Normand, M. B. Landrum, and R. A. Rosenheck. 2002. "Longitudinal Profiles of Health Care Providers." *Statistics in Medicine* 21: 1067–88.
- Carey, J. S., J. P. Parker, C. Brandeau, and Z. Li. 2008. "The 'Occasional Open Heart Surgeon' Revisited." *Journal of Thoracic and Cardiovascular Surgery* 135: 1254–60.
- Cohen, W. M., and D. A. Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Sciences Quarterly* 35 (1): 128–52.
- Denvir, M. A., A. J. Lee, J. Rysdale, R. J. Prescott, H. Eteiba, A. Walker, P. Mankad, and I. R. Starkey. 2006. "Variations in Clinical Decision-Making between Cardiologists and Cardiac Surgeons; a Case for Management by Multidisciplinary Teams?" *Journal of Cardiothoracic Surgery* 1 (1): 3–31.
- Edmondson, A. C., A. B. Winslow, R. M. Bohmer, and G. P. Pisano. 2003. "Learning How and Learning What: Effects of Tacit and Codified Knowledge on Performance Improvement Following Technology Adoption." *Decision Sciences* 34 (2): 197–223.
- Ferraris, V. A., F. H. Edwards, D. M. Shahian, and S. P. Ferraris. 2008. "Chapter 7: Risk Stratification and Comorbidity." In *Cardiac Surgery in the Adult*, edited by L. Cohn, pp. 199–246. New York: McGraw-Hill.
- Gawande, A. 2007. "The Checklist: If Something So Simple Can Transform Intensive Care, What Else Can It Do?" *The New Yorker* 10: 86–101.
- Gaynor, M., H. Seider, and W. B. Vogt. 2005. "The Volume–Outcome Effect, Scale Economies and Learning-by-Doing." *American Economic Review* 2: 243–7.
- Hales, B. M., and P. J. Pronovost. 2006. "The Checklist—A Tool for Error Management and Performance Tracking." *Journal of Critical Care* 21 (3): 231–5.
- Hannan, E. L., C. Wu, T. J. Ryan, T. Bennett, A. T. Culliford, J. P. Gold, A. Hartman, O. W. Isom, R. H. Jones, B. McNeil, E. A. Rose, and V. A. Subramanian. 2003. "Do Hospitals and Surgeons with Higher Coronary Artery Bypass Graft Surgery Volumes Still Have Lower Risk-Adjusted Mortality Rates?" *Circulation* 108: 795–801.
- Hao, L. 2008. "Assortative Matching." In *The New Palgrave Dictionary of Economics*. 2d Edition, edited by S. N. Durlauf and L. E. Blume. New York: Palgrave Macmillan. "The New Palgrave Dictionary of Economics Online" [accessed April 17, 2010], doi:10.1057/9780230226203.0066. Available at http://www.dictionaryofeconomics.com/article?id=pde2008_A000214.

- Hemingway, H., A. M. Crook, G. Feder, S. Banerjee, J. R. Dawson, P. Magee, S. Philpott, J. Sanders, A. Wood, and A. D. Timmis. 2001. "Underuse of Coronary Revascularization Procedures in Patients Considered Appropriate Candidates for Revascularization." *New England Journal of Medicine* 344 (9): 645–54.
- Hirth, R. A., M. N. Turenne, J. R. C. Wheeler, Q. Pan, Y. Ma, and J. M. Messana. 2009. "Provider Monitoring and Pay for Performance When Multiple Providers Affect Outcomes: An Application to Renal Dialysis." *Health Services Research* 44 (5): 1585–602.
- Huckman, R. 2003. "The Utilization of Competing Technologies within the Firm: Evidence from Cardiac Procedures." *Management Science* 49 (5): 599–617.
- Huckman, R., and G. Pisano. 2006. "The Firm-Specificity of Individual Performance: Evidence from Cardiac Surgery." *Management Science* 52 (4): 473–88.
- Isaac, T., J. S. Weissman, R. B. Davis, M. Massagli, A. Cyrulik, D. Z. Sands, and S. N. Weingart. 2009. "Overrides of Medication Alerts in Ambulatory Care." *Archives of Internal Medicine* 169 (3): 305–11.
- Kee, F., P. McDonald, J. R. Kirwan, C. C. Patterson, and A. H. G. Love. 1998. "Urgency and Priority for Cardiac Surgery: A Clinical Judgment Analysis." *British Medical Journal* 316 (7135): 925–9.
- Kramer, A. A., and J. E. Zimmerman. 2007. "Assessing the Calibration of Mortality Benchmarks in Critical Care: The Hosmer-Lemeshow Test Revisited." *Critical Care Medicine* 35 (9): 2052–6.
- Luft, H. S., S. S. Hunt, and S. B. Maerki. 1987. "The Volume–Outcome Relationship: Practice-Makes-Perfect or Selective-Referral Patterns?" *Health Services Research* 22 (2): 157–82.
- Marcin, J. P., L. Zhongmin, R. L. Kravitz, J. J. Dai, D. M. Rocke, and P. S. Romano. 2008. "The CABG Surgery Volume–Outcome Relationship: Temporal Trends and Selection Effects in California, 1998–2004." *Health Services Research* 43 (1): 174–92.
- Miyata, H., N. Motomura, Y. Ueda, H. Matsuda, and S. Takamoto. 2008. "Effect of Procedural Volume on Outcome of Coronary Artery Bypass Graft Surgery in Japan: Implication Toward Public Reporting and Minimal Volume Standards." *Journal of Thoracic and Cardiovascular Surgery* 135: 1306–12.
- Neuhaus, J. M., and J. D. Kalbfleisch. 1998. "Between- and Within-Cluster Covariate Effects in the Analysis of Clustered Data." *Biometrics* 54 (2): 638–45.
- O'Connor, G. T., S. K. Plume, E. M. Olmstead, J. R. Morton, C. T. Maloney, W. C. Nugent, F. Hernandez, Jr., R. Clough, B. J. Leavitt, L. H. Coffin, C. A. Marrin, D. Wennberg, J. D. Birkmeyer, D. C. Charlesworth, D. J. Malenka, H. B. Quinton, and J. F. Kasper. 1996. "A Regional Intervention to Improve the Hospital Mortality Associated with Coronary Artery Bypass Graft Surgery." *Journal of the American Medical Association* 275 (11): 877–8.
- Pennings, J. M., and F. C. Wezel. 2007. *Human Capital, Inter-Firm Mobility and Organizational Evolution*. Cheltenham, England: Edward Elgar Publishing.
- Peterson, E. D., L. P. Coombs, E. R. DeLong, C. K. Haan, and T. B. Ferguson. 2004. "Procedural Volume as a Marker of Quality for CABG Surgery." *Journal of the American Medical Association* 291: 195–201.

- Pietz, K., K. J. O'Malley, M. Byrne, J. Soucek, N. Petersen, and C. M. Ashton. 2002. “Physician Level Variation in Practice Patterns in the VA Healthcare System.” *Health Services and Outcomes Research Methodology* 3: 95–106.
- Pisano, G. P., M. J. Bohmer, and A. C. Edmondson. 2001. “Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery.” *Management Science* 47: 752–68.
- Rabe-Hesketh, S., and A. Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata*. 2d Edition. College Station, TX: Stata Press.
- Reagans, R., L. Argote, and D. Brooks. 2005. “Individual Experience and Experience Working Together.” *Management Science* 51: 869–81.
- Schofield, P. M. 2003. “Indications for Percutaneous and Surgical Revascularisation: How Far Does the Evidence Base Guide Us?” *Heart* 89 (5): 565–70.
- Shahian, D. M. 2004. “Improving Cardiac Surgery Quality—Volume, Outcome, Process?” *Journal of the American Medical Association* 291 (2): 246–8.
- Shahian, D. M., F. H. Edwards, V. A. Ferraris, C. K. Haan, J. B. Rich, S. L.-T. Normand, E. R. DeLong, S. M. O'Brien, C. M. Shewan, R. S. Dokholyan, and E. D. Peterson. 2007. “Quality Measurement in Adult Cardiac Surgery: Part 1—Conceptual Framework and Measure Selection.” *Annals of Thoracic Surgery* 83 (4): S3–12.
- Shahian, D. M., and S. L.-T. Normand. 2003. “The Volume–Outcome Relationship: From Luft to Leapfrog.” *Annals of Thoracic Surgery* 75: 1048–5.
- . 2008. “Low-Volume Coronary Artery Bypass Surgery: Measuring and Optimizing Performance.” *Journal of Thoracic and Cardiovascular Surgery* 135: 1202–9.
- Torchiana, D. F., and G. S. Meyer. 2005. “Use of Administrative Data for Clinical Quality Measurement.” *Journal of Thoracic and Cardiovascular Surgery* 129 (6): 1223–5.
- Tsai, A. C., M. Votruba, J. F. P. Bridges, and R. D. Cebul. 2006. “Overcoming Bias in Estimating the Volume–Outcome Relationship.” *Health Services Research* 41 (1): 252–64.
- Von Hippel, E. 1994. “‘Sticky Information’ and the Locus of Problem Solving: Implications for Innovation.” *Management Science* 40 (4): 429–3.
- Wennberg, D. E. 1999. *Dartmouth Atlas of Cardiovascular Health Care Working Group. The Dartmouth Atlas of Cardiovascular Health Care*. Chicago, IL: AHA Press.
- Young, G. J., M. P. Charns, K. Desai, S. F. Khuri, M. G. Forbes, W. Henderson, and J. Daley. 1998. “Patterns of Coordination and Clinical Outcomes: A Study of Surgical Services.” *Health Services Research* 33 (5): 1211–36.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Table SA1. Hospital Volumes by Type and Year.

Table SA2. Surgeon Volumes by Type and Year.

Table SA3. Selection of Observations and Missing Data.

Table SA4. Stata Code and Regression Output for All Models Presented in the Text.

Table SA5. Risk Model Discrimination Analysis.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.