Refinements in Evaluating Minimum Surgery Volume Standards

PI: Amresh D. Hanchate, PhD, Boston University School of Medicine

Co-Investigators:
Arlene S. Ash, PhD, Boston University School of Medicine
Therese Stukel, PhD, Institute of Clinical Evaluative Sciences, Toronto, Canada

Dates of project: 9/30/2005 – 9/29/2007

Project Officer: Yen-Pin Chiang

Grant number: 1R03HS015617-01A1

We gratefully acknowledge the financial support from the aforementioned AHRQ grant in conducting the project research.

Abstract

***Objective*:** For many surgical procedures, apparent volume-outcome relationships may reflect differences in patient risk-profiles as well as quality of care. As some important patient profile differences may be unobserved, we use fixed effects (FE) regression to estimate the relationship between operative mortality and surgeon and hospital volumes, and we compare this method with the more commonly used random effects (RE) regression approach.
Data Sources: The 1998 and 1999 Medicare Inpatient and Denominator files for Medicare Fee for Service enrollees aged 65 to 99.
Study Design: Operative mortality rates are estimated for different surgeon and hospital volume tertiles (high, medium, low) using FE and RE regression methods, adjusted for patient demographics and morbidities.
Data Collection: The data were collected by the Centers for Medicare and Medicaid Services (CMS).
Principal Findings: FE regression estimates that lowest volume tertile hospitals have 1.4, and lowest volume tertile surgeons have 1.6, additional operative deaths (for every 100 CABG surgeries) compared to their highest volume tertile counterparts. The corresponding RE estimates are 0.5 and 1.4, respectively. The substantially higher FE hospital volume effect compared to RE indicates the presence of unobserved "protective" characteristics in lower-volume providers, including a less complicated patient profile.
Conclusions: Lower hospital and surgeon volumes are associated with substantially higher excess operative mortality from CABG surgeries than previously estimated.

Keywords: hospital volume, surgeon volume, fixed effects, random effects

PURPOSE

Findings that higher surgery volumes are associated with lower operative deaths have sparked off a widespread response from consumer advocacy groups, health insurance coalitions (Leapfrog), and state agencies, all aimed at transferring patients away from low-volume providers as a key mechanism to reduce risk of operative death. Minimum volume thresholds for hospitals are already in practice in determining referrals for selected high-risk surgeries. One example is the Leapfrog minimum hospital-level threshold of 450 coronary-artery bypass graft (CABG) surgeries per year.

In a seminal study that compared the impact of surgery volumes and hospital volumes side-by-side, Birkmeyer et al (1) found that, for CABG surgeries, the volume of hospital had no significant protective effect on risk of operative death. Instead, it was the surgeon volume that had a large protective effect. That is, low-volume surgeons at all hospitals have the same higher risk of operative death, whereas high-volume surgeons at all hospitals have the same lower risk of operative death. If confirmed, these findings call to question the logic for hospital-level volume thresholds. The relative role of the two volumes varied with across different types of the 14 high-risk surgical procedures evaluated in the study.

Though Birkmeyer et al (1) have been very insightful and influential in raising new issues in an established strand of literature, the statistical estimation model used rests on assumptions that may be too strong.

The primary objective of this study is to re-examine the data from Birkmeyer et al (1) using an alternative estimation model that is broader in scope and better suited for delineating the operative mortality effect of surgeon volume from that of hospital volume. Our purpose is to evaluate if the findings of Birkmeyer et al (1) robust to these adjustments. We have first examined this for CABG surgery – these results have been reported in a manuscript submitted for publication. A second manuscript relating to other surgical procedures is under preparation. The details below are described in the context of CABG surgeries.

BACKGROUND

An overwhelmingly large proportion of studies to date indicate that hospitals and surgeons who perform more surgeries have lower operative mortality (2, 3). To what extent, then, can we infer that higher-volume providers offer better quality care? At the provider level, lower operative mortality can result from better quality of care or from having more patients with fewer medical complications. Such variation of patients across providers can result from selective physician referral, patient choice, or strategic selection by providers ("cherry picking"). An analyst using administrative or clinical data can identify and control for some of these patient profile differences using available clinical and demographic information. But there may still be other patient characteristics, largely unobserved by the analyst, that potentially influence operative mortality. We use a fixed effects (FE) regression modified to take advantage of the dual clustering of patients among surgeons and hospitals, to estimate the association between provider volume and operative mortality in a way that adjusts even for unobserved provider-level differences in patient characteristics. Results are then compared with those from random effects (RE) regressions, the most commonly used approach in the literature.

Though the present study is limited to CABG surgeries, the significance of this methodological issue should be viewed in the context of the far-reaching impact of this voluminous literature, dating back over two decades (4). As indicated by a recent survey, interest in the volume-mortality relationship spans a wide range of surgeries (predominantly high-risk cardiovascular surgeries and cancer resections) as well as nonsurgical care (such as inpatient treatment for pneumonia or HIV-positive patients) (2, 5). The findings from this literature have attracted the attention of a variety of interest groups – consumer advocacy groups, health insurance coalitions, and state agencies seeking to reduce costs and improve quality by enforcing protocols of proven efficacy. Strategies include regionalizing selected surgeries, publishing provider report cards, and recommending provider minimum volumes for specified surgeries (6-10).

Along with the growth of the volume-mortality literature has come a better understanding of its vulnerabilities, especially given that virtually all such studies are based on observational data, usually administrative or clinical databases. An enduring potential weakness arises from unobserved differences among patients seen by different providers. The multitude of processes that connect patients with surgeons and hospitals are not random and are known to result in systematic differences, often large, in the patients treated by different surgeons at different hospitals. Sicker patients are likely to gain more from higher-quality care and thus may be more prevalent among the patients of providers perceived to have higher quality. This matching may result from physician referrals or from patients' choice based on available information (including report cards).

Another sorting process involves a different kind of response to provider report cards: providers avoiding sicker patients ('cherry picking'), possibly seeking a more favorable operative mortality record (11). Recent evidence also points to systematic differences by hospitals in socioeconomic patient profiles – in particular, of significant clustering of Black and other minority patients in relatively few hospitals (12, 13). Not all the important differences among patients are observed, even in detailed clinical databases (11, 14). To the extent that these unobserved factors significantly affect operative mortality, then traditional comparisons of operative mortality rates across all surgeons and hospitals risk mistaking differences in patient severity with differences in the quality of care. For instance, if higher-volume hospitals attract disproportionately larger number of sicker patients and some important illness characteristics are not observed, then RE regression is likely to overestimate adjusted mortality and underestimate quality of care for the surgeons in higher-volume hospitals. An attractive alternative is to limit comparisons of surgeons within each hospital, thereby sweeping out unobserved patient characteristics across hospitals. This is the basic logic behind the FE regression.

Given the clustering of patients at the surgeon and hospital levels, the advantage of FE regression is in exploiting the within-provider variation in operative mortality to estimate the volume-outcome relationships, thereby making it robust to systematic differences in unobserved characteristics at provider level. In contrast, the RE regression, the standard workhorse in this literature, is based on the assumption that there are no systematic differences across providers in unobserved patient characteristics. The only existing study using FE regression that we know of examined associations of longitudinal changes in hospital volume on two patient outcomes (length of stay and inpatient mortality) from hip fracture surgery (15). They found that the apparent protective effect of higher volume from RE regression disappeared when adjusted for hospital fixed effects.

We use a readily available data set previously used to examine the association of operative mortality with hospital and surgeon volumes. To control for systematic provider-level differences in unobserved patient as well as provider characteristics, we use a FE regression approach modified to take advantage of the nature of surgeon- and hospital-level clustering of patients. Two separate FE regressions are estimated, one to estimate the surgeon volume effect (i.e., association of surgeon volume and operative mortality) and another to estimate the hospital volume effect. As most hospitals in our data had two or more surgeons, to estimate surgeon volume effect, unobserved differences in patients across hospitals are controlled for by only comparing operative mortality of surgeons within same hospital. To estimate the hospital volume effect, we take advantage of the fact that surgeons are not nested in hospitals – a large proportion (51 percent) of surgeries are performed by surgeons who operated at two or more hospitals. This enables us to compare operative mortality across hospitals of patients operated by the same surgeon – thereby adjusting for systematic unobserved patient characteristics at the surgeon level.

By relying only on within-cluster comparisons, the FE approach offers a better approximation of the operative mortality differences arising from differences in quality of care indicators – including caregiver skill, experience, and pre- and postoperative processes of care. This herein is referred to as the *quality of care component* of the operative mortality differences by provider volume.

Based on this, we can also estimate the *unobserved factors component* – the residual mortality differences by provider volume that may be attributed to unobserved characteristics, including systematic unobserved patient differences across providers. We compare the estimates of this FE decomposition with the overall single estimate from a parallel RE regression. That is, does the sum of the two FE components equal the estimate from RE regression?

METHODS

Data

We use an analytic data set of patient-level CABG surgery mortality outcomes and covariates previously used to examine volume-outcome relationships, adopting all the variable definitions from that study (1, 16). Briefly, using 100 percent of acute care hospitalization discharge data for Medicare fee-for-service beneficiaries in 1998 and 1999, admissions in which CABG surgery was performed for persons aged 65 to 99 are included (thereby excluding a small number of CABGs performed on younger patients with disability or End-Stage Renal Disease). Discharges that also involved a valve replacement were excluded. The Institutional Review Board at Boston University School of Medicine approved the study protocol.

To identify the operating surgeon for each CABG, the unique provider identification number in the “primary operator” field in the Medicare Inpatient file was used. In 6% of procedures the provider identification numbers were invalid and therefore excluded. In addition, only CABGs performed by self-designated cardiothoracic surgeons were selected, to avoid cardiologists being wrongly identified as surgeons. This results in an additional 13% of the records being excluded, leading to a study sample of 220,592 patients.

Analytic cohorts

Patients in the data are clustered at the level of surgeons and hospitals. The FE approach to estimating the effect of surgeon volume consists of comparing surgeons in each hospital – thereby requiring at least two surgeons in every hospital. This cohort, herein called the *Within Hospital Cohort*, is obtained by excluding 60 hospitals (of 958), because only one surgeon operated there, and 2,802 patients (of 220,592).

An analogous FE approach is used to estimate the effect of hospital volume. Here, we limit our analysis to outcome for patients of surgeons who operate at two or more hospitals (“splitters”) so that we can compare outcome of patients from the same surgeon but at different hospitals. This cohort, called the *Within Surgeon Cohort*, retains 44% of all surgeons, 79% of the hospitals, and 51% of all patients.

Outcome measure and covariates

Operative mortality for a patient is defined as death within 30 days of the procedure or before hospital discharge. Surgeon and hospital volumes are defined as the total number of CABG surgeries performed in a year, including Medicare as well as other payer patients. These are estimated for individual surgeons and hospitals by scaling up Medicare FFS volume to reflect total volume – the scale up multipliers are based on the proportion of all CABG patients who are Medicare FFS beneficiaries, obtained from 1997 Nationwide Inpatient Sample (NIS) and urban/rural location. Both surgeon and hospitals are categorized into patient-level tertiles – using 101 (33^{rd} percentile) and 162

(66th percentile) surgeries per year as the cutoffs for surgeon volume and 314 and 628 for hospital volume. Patient covariates include age, sex, race, the Charlson comorbidity score, and an area-level income measure (mean income from Social Security for the patient's residence ZIP code). The Charlson score is based on ICD-9-CM diagnosis codes from the index admission as well as any other inpatient admissions in the preceding 6 months, excluding primary indicators for the surgical procedure or postoperative complications. Hospital characteristics adjusted for are teaching status and ownership (not-for-profit, government, and for-profit).

Estimation

Our interest is in estimating the relationship between operative mortality and both surgeon and hospital volumes. A general regression notation that encompasses the FE and RE approaches is as follows.

$$OM_{psh} = \alpha * PC_{psh} + \beta_s * SV_{sh} + \beta_h * HV_{sh} + \gamma * HC_{sh} + u_s + v_h + e_{sph}$$

OM denotes operative mortality (with values 0 and 1, denoting survival and death respectively) of patient *p* operated by surgeon *s* in hospital *h*. The covariates are grouped as patient characteristics (*PC*), surgeon volume (*SV*), hospital volume (*HV*), and other hospital characteristics (*HC*). Unobserved cluster effects at the surgeon and hospital levels are denoted by *u* and *v*, respectively. Finally, *e* denotes the effect of unobserved patient characteristics. Both FE and RE regression models are estimated as linear regressions[1]. In both models, *e* is specified to be independent and identically (normal) distributed random variable with mean 0. Because all regression covariates are categorical groups, all the regression coefficients are interpreted as excess rates in operative mortality compared to the reference category.

RE estimates are obtained from a three-tiered hierarchical regression wherein surgeons are treated as being nested with in hospitals – with surgeons operating at two or more hospitals treated as distinct surgeons, using their combined volume across all hospitals (17). This regression is estimated for both the analytic cohorts (within hospital and within surgeon) using the *xtmixed* procedure in Stata 9.2 (18).

The FE estimates are obtained from two separate regressions, each estimating the volume effects for the two provider types, surgeons, and hospitals. To estimate the effect of surgeon volume, we limit comparisons of surgeons within the same hospital by the following transformation of equation (1), wherein the outcome measure as well as all the covariates are expressed in terms of within-hospital mean differences (19).

$$(OM_{psh} - \overline{OM}_h) = \alpha * (PC_{psh} - \overline{PC}_h) + \beta_s * (SV_{sh} - \overline{SV}_h) + (u_s - \bar{u}_h) + (e_{sph} - \bar{e}_h) \quad (2)$$

Each variable transformation involves differencing cluster level means – for instance, $\overline{OM}_h$ denotes the mean operative mortality among all patients at hospital *h*, and $\bar{u}_h$ denotes the mean unobserved surgeon effects (u_s) across all the surgeons in hospital *h*. Note that, with all hospital measures (*HV*, *HC*, and *v*) eliminated, the resulting equation (2) has a two-tier structure (patients and surgeons). This transformed equation is estimated as a RE linear regression model using the within-hospital cohort.

[1] Fixed effects logistic regression (also known as conditional logistic regression) requires at least one decedent and survivor from each fixed level (surgeon, hospital) (Chamberlain 1980). As 507 of the 2,772 surgeons have no decedents, we have instead followed previous studies (Tsai 2006) and chosen to use the linear probability specification that has the advantage of retaining data from all surgeons and hospitals.

This in effect is a partial or quasi FE approach. The RE regression of (2) assumes that the mean differenced unobserved surgeon effect, $(u_s - \bar{u}_h)$, is uncorrelated with other covariates and error term, all of which are mean differenced – implying that, for instance, within-hospital differences around mean in volume across surgeons, $(SV_{sk} - \overline{SV_h})$, and is uncorrelated with within-hospital differences around mean in patient severity across surgeons, $(u_s - \bar{u}_h)$. However, this permits cluster-level measures (say, $\overline{SV}$ and $\bar{u}_h$) to be correlated.

Analogously, to estimate the effect of hospital volume, we use the within-surgeon cohort with only patients of surgeons who operated at two or more hospitals – thereby permitting comparison of outcomes of patients across hospitals holding the surgeon characteristics the same. The corresponding transformation of (1) is

$$(OM_{psh} - \overline{OM_s}) = \alpha * (PC_{psh} - \overline{PC_s}) + \beta_h * (HV_{sh} - \overline{HV_s}) + \gamma (HC_{sh} - \overline{HC_s}) + (v_h - \bar{v}_s) + (e_{psh} - \bar{e}_s) \quad (3)$$

$\overline{OM_s}$ denotes the mean operative mortality for surgeon *s* patients across all the hospitals. Note that, here, surgeon characteristics (*SV* and *u*) are eliminated and the resulting structure is two tiered (patients and hospitals). This equation is estimated as a RE linear regression model using the within-surgeon cohort.

The FE hospital volume effect is analogously estimated by limiting comparisons of operative mortality of patients with the same surgeon but who operates at different hospitals (i.e., the within surgeon cohort). Here, transformation involves differencing surgeon-level averages, followed by a two-tiered RE regression involving hospital unobserved cluster effect. Additional technical details are elaborated in the Appendix. The *quality of care component* of the volume effect on operative mortality is measured by β_h for hospital volume and β_s for surgeon volume, both estimated from the FE regressions.

FE and RE approaches differ principally in the specification of cluster effects u_s and v_h. Note that v_h represents the operative mortality at hospital *h* resulting from unobserved factors that vary systematically across hospitals, and u_s is the analog for unobserved differences across surgeons. Therefore, systematic differences in unobserved patient characteristics, if any, are captured by either u_s or v_h. The RE estimation assumes that u_s and v_h are uncorrelated with the model covariates (*PC, SV, HV,* and *HC*) as well as residual *e* – in particular, this implies that there are no systematic differences in unobserved patient characteristics across providers. Violation of this assumption results in biased estimates – β_h and β_s from RE no longer reflect the *quality of care component*. In contrast, the FE approach makes a weaker assumption, permitting this correlation at both the cluster levels (surgeons and hospitals) but assuming that within-cluster differences across measures are uncorrelated. This implies that, even if provider volume were correlated with unobserved patient severity, β_h and β_s from FE remain unbiased estimates of *quality of care component*. Therefore, the greater the differences in FE and RE estimates of β_h and β_s, the greater the influence of unobserved factor differences by providers.

The overall *unobserved factors component* is measured by the residual mortality – that is, the combined effect of unobserved hospital and surgeon factors (u+v). This is obtained by substituting the FE covariate estimates in equation (1) to obtain

$$OM_{psh} - \alpha * PC_{psh} + \beta_s * SV_{sh} + \beta_h * HV_{sh} + \gamma * HC_{sh} = u_s + v_h + e_{sph} \qquad (4)$$

We report the average of this measure by each provider volume.

Therefore, we compare RE estimates of β_h and β_s not only against those from FE but also against the sum of *quality of care* and *unobserved factors components* from FE regressions.

RESULTS

CABG patients, surgeons, and hospitals

Table 1 provides sample characteristics of the entire study population (All column) as well as the two subsets – the within hospital and the within surgeon cohorts. The entire study population consists of 220,592 patients operated on by 2,772 surgeons in 958 hospitals during 1998 and 1999. The overall operative mortality was 50.5 per 1,000 surgeries. The average annual volume for surgeons was 85 and that for hospitals was 297 – note that these volumes refer not only to Medicare patients but to all patients. As mentioned earlier, 44% of surgeons operated at two or more hospitals (i.e., splitters), and 94% of hospitals had two or more surgeons.

All analyses are performed in terms of patient-level tertiles of hospital and surgeon volumes. Among the 2,772 surgeons, 377 were in the top tertile, with annual volume ranging from 162 to 567 surgeries, and 1,783 were in the bottom tertile, with a volume of less than 101 surgeries. Of the 958 hospitals, 101 are in the top tertile (again of patients) with an annual volume of at least 628 surgeries; 644 hospitals are in the bottom tertile with a volume of less than 314 surgeries. Patients in high/low-volume hospitals (i.e., highest/lowest-volume tertile) were more likely to also have a high/low-volume surgeons, and vice versa. Of the patients at high-volume hospitals, half were operated on by high-volume surgeons, and about 15%, by low-volume surgeons; for the patients at low-volume hospitals, half were operated on by low-volume surgeons, and about 15%, by high-volume surgeons. A similar pattern was observed for the converse distribution of patients at low- and high-volume surgeons across low- and high-volume hospitals.

The analytic cohorts

All summary figures in Table 1 for the within hospital cohort are virtually identical to that for the entire sample. However, the within surgeon cohort, containing surgeons operating at two or more hospitals, shows differences in provider profiles – surgeon volumes are larger (by 10%), hospital volumes are smaller (by 17%), and fewer hospitals are teaching (by 15%). More importantly, the patient profile appears to be no different compared to the overall population, and operative mortality rates and patient characteristics are basically identical.

Table 2 gives the unadjusted operative mortality rates (per 1,000 CABG surgeries) for different strata cross-classified by provider volume for both analytic cohorts. Lower-volume hospitals had 11 more operative deaths, whereas lower-volume surgeons had 14 more operative deaths compared with their high-volume counterparts. Higher mortality is associated with older age, female gender, black race, higher Charlson score, and emergent admission. With respect to provider volume, two patterns emerged –

i) the magnitude of difference between high- and low-volume providers (either hospitals or surgeons) is constant across most strata, and ii) the two analytic cohorts have very similar operative mortality rates for the same strata.

Estimates of surgeon and hospital volume effects

Table 3 presents the main regression estimates from both FE and RE approaches. Column 1 presents the RE estimates of operative mortality rates (%) associated with surgeon and hospital volumes using the within hospital cohort. For comparison this model is also estimated for the within surgeon cohort (column 2). Columns 3 and 4 give the corresponding estimates from the FE approach. Note that both the RE estimates are similar across all patient and hospital characteristics, although the hospital volume effects are smaller in column (2) but not (statistically) significantly different; the column (1) figure will be used for RE estimates herein. Using patients at high-volume providers as the reference cohort, those treated by low- and medium-volume hospitals had 0.45 more operative deaths per 100 CABG surgeries (95% CI=[0.0008, 0.89]), and those treated by low-volume surgeons had 1.41 more operative deaths (95% CI =[1.08, 1.74]). The FE approach estimates that, compared to their high-volume provider counterparts, patients treated by low-volume hospitals had 1.36 more operative deaths (95% CI=[0.34, 2.37]), and those by low-volume surgeons had 1.56 more operative deaths (95% CI=[0.67, 2.46]).

Estimates of other covariates that have statistically significant association with operative mortality – gender and age composition, Charlson score, and admission type – have very similar estimates across all four regressions (columns 1 to 4). The teaching status of the hospital and the mean Social Security income in a patient's residence zip code were not associated with operative mortality in any of the models. However, for-profit hospitals had higher operative mortality in the RE models but not in the FE model.

As described in the Methods section, FE regression is used to decompose the total volume effect into a *quality of care* component (obtained from FE regression coefficients in Table 3) and an *unobserved factor* component (hospital or surgeon-level mean of FE regression residuals). These are presented in the first two columns of Table 4, followed by their sum in column 3. As the *unobserved factor* component (column 2) gives the portion of the observed operative mortality not explained by the regression variables, negative values indicate that observed mortality was less than expected by the FE model. We see that low-volume hospitals have 0.84 fewer deaths than expected (using highest-volume tertile as the reference); and middle- (tertile) volume hospitals have 0.40 fewer deaths – resulting from unobserved factors, including a less complicated patient profile. Column 3 gives the sum of the two FE components. Combining the two components – 1.36 additional deaths from the quality component and 0.84 fewer deaths due to other unobserved factors, lowest-tertile hospitals have a total (net) excess mortality of 0.52 deaths (per 100 CABG surgeries) – as indicated in column 3. This figure is similar to the direct estimate of 0.45 excess deaths from RE regression (column 4). A similar pattern is true for medium-volume hospitals, with a net FE effect of 0.40 and a direct RE estimate is 0.39 excess deaths, thereby suggesting that RE hospital volume effects approximate the sum of the quality and unobserved factor FE components. But this comparison does not hold for the surgeon volume comparison – particularly for the low-volume surgeons. The average *unobserved factors component* for low-volume surgeons is –5.8, thereby leading to the total FE estimate of 9.8 excess mortality rate – much lower than the RE direct estimate of 14.1.

Sensitivity Analyses

An important step in the estimation process is the use of the two distinct subsets (cohorts) of the overall data. In particular, the within surgeon cohort excludes 49% of the patients and 44% of surgeons. We performed a number of robustness and sensitivity checks to validate the estimates obtained. First, to assess the robustness of the FE hospital effects estimates, we re-estimated the regression using bootstrapping (1,000 replications), randomly dropping 10% of the surgeons (and their patients). The results do not change – the average excess mortality rate for low-volume hospitals is 1.4 (95% CI=[0.89, 1.86]) and for medium-volume hospitals is 0.8 (95% CI=[0.42, 1.29]).

Second, we address the difference in hospital volume effects estimated by the two RE regressions (columns 1 and 2 of Table 3). For instance, for low-volume hospitals, the excess mortality estimate is 0.45 using the within hospital cohort (column 1) but it is 0.09 using the within surgeon cohort (column 2). Note that, although the within-hospital cohort includes virtually all the study data, the within-surgeon cohort excludes all patients treated by 44% of surgeons who operated at more than one hospital. It is unclear if this difference is indicative of systematic difference in the within-surgeon cohort subset or indicative of lack of robustness of the estimate. To evaluate, this we obtained 100 different subsamples of the overall data by excluding all patients from 44% of surgeons randomly selected – that is, exclusion of surgeons was not based on whether or not they operated at more than one hospital. The mean (from the 100 regressions) of low hospital volume effect is 0.39, and the range is –0.17 to 0.88 – thereby suggesting that the aforementioned difference (betweens columns 1 and 2 of Table 3) may be due to the lack of robustness of the estimate and not necessarily indicative of systematic differences in patients in the two cohorts.

To further ascertain if the patients included in the within surgeon cohort are systematically different from those who are excluded, and in the spirit of marginal propensity score estimation, we performed a logistic regression with inclusion/exclusion as the outcome and all patient factors plus operative mortality as the covariates. We find that this model has poor discrimination – with only 51% of patients correctly classified (area under ROC was 51.3), thereby indicating little systematic difference between the included and excluded patients.

Finally, to examine if the results are sensitive to using distinct cohorts for the two FE regressions, we identified a subsample that simultaneously met the two criteria identifying each cohort (hospitals with at least two surgeons, and surgeons who operated at two or more hospitals); this subsample has 104,340 patients, 596 hospitals, and 1,126 surgeons. The RE and both the FE regressions estimated on this common data yielded virtually the same results as those in Table 3 – in particular, a much higher estimate of excess deaths associated with low hospital volume from FE regression (1.21 deaths) than from RE regression (0.66 deaths).

DISCUSSION

Are lower operative mortality rates among higher-volume surgeons and hospitals a signal of better quality of care, or is this association confounded by patient profile differences across providers?

Providers may differ in the proportion of complicated patients they treat, and not all these complications are adequately identifiable in administrative or clinical data. To better adjust for such differences, we used fixed effects (FE) regression methods to obtain estimates based only on within cluster comparisons. To avoid comparisons across hospitals, surgeons within same hospital were compared with each other in estimating the surgeon volume effect; to estimate the hospital volume effect, we compared operative mortality across hospitals for patients treated by the same surgeon. By overcoming the potential confounding from unobserved patient differences, this approach better approximates the volume effects associated with *quality of care* differences across providers. This regression approach indicates that, compared to hospitals that perform at least 628 CABG surgeries a year, those that perform fewer than 314 a year have an excess operative mortality rate of 1.36 deaths (per 100 surgeries), and those that perform between 314 and 628 surgeries have an excess mortality of 0.8. Furthermore, compared to surgeons who perform at least 162 CABG surgeries a year, those who perform fewer than 101 surgeries have an excess operative mortality rate of 1.56.

Comparing FE and RE regression estimates indicate that they differ significantly with respect to the effects of hospital volume but not surgeon volume. Specifically, FE regression estimates that excess deaths associated with low-volume hospitals (1.36 excess deaths per 100 CABG surgeries) is much higher than that estimated by RE regression (0.52). Recall that the FE estimate is based on comparing operative mortality of same surgeon across hospitals – implying that, even after adjusting for surgeon-level factors, there are significant differences across hospitals in unobserved factors. On the other hand, FE and RE estimates of excess deaths associated with low surgeon volume are similar (1.56 and 1.41, respectively) – because FE estimates are based on within-hospital comparison of surgeons, similarity between FE and RE estimates implies that, once unobserved factors at the hospital level are adjusted for (i.e., within each hospital), unobserved factors across surgeons by volume do not have significant effect on operative mortality. That is, within each hospital, there are no systematic differences in unobserved patient characteristics across surgeons – consequently, the surgeon volume effects from RE and FE are very similar.

The FE approach also enables estimation of the effect of unobserved characteristics on operative mortality (*unobserved factors component)* at the provider level. This captures factors that are important determinants of operative mortality but not observed in the data. Averaging this component at provider volume level, we indeed find large differences in this component. Low-volume hospitals and surgeons have much lower operative mortality than expected by the FE regression model. Though the regression model predicts low-volume hospitals to have an excess operative mortality rate of 1.36, the observed excess mortality rate is 0.52, leading to a large difference accounted for by unobserved factors. Similarly, although the low-volume surgeons have an expected excess mortality rate of 1.56, the observed excess mortality rate is 0.98. Both of these indicate the presence of large "protective" factors in low-volume hospitals and providers. Note that process-of-care differences across providers that vary with volume are already captured explicitly in the regression, but other process differences (that affect quality) could contribute to this unobserved component.

Based on growing evidence from other studies, a plausible "protective" factor may be unobserved patient severity or complications, implying that high-volume hospitals and surgeons have more complicated patients compared with their low-volume counterparts. Although this implication of a more complicated patient profile in higher-volume hospitals contradicts some previous findings (2), a number of recent studies corroborate it. The strongest evidence comes from studies of the impact of mandatory annual reporting of risk-adjusted mortality from CABG surgery for each hospital and surgeon in New York and Pennsylvania since 1992. One study surveyed a sample of cardiologists and cardiac surgeons from the two states, finding that 59% of the cardiologists "reported increased difficulty in finding surgeons willing to perform CABG surgery in severely ill patients who required it, and 63% of the cardiac surgeons reported that they were less willing to operate on such patients" (20). Another large study based on the majority of all CABG surgeries performed among Medicare population between 1987 and 1994 found that "report cards led to increased sorting of patients to providers on the basis of the severity of their illness … with those two states' teaching hospitals picking up an increasing share of patients with more severe illness" (11). Because physician referral has been found to be the most important determinant of provider choice, it is likely that more complicated patients are referred to teaching and other high-volume providers, as they are expected to gain more than those with fewer complications (21). Despite this supportive evidence in the literature, we acknowledge that the present study provides not direct evidence of unobserved patient profile differences by provider volume. There may be other unobserved factors, including at provider levels, not adequately captured by the methods used here.

This study also has implications for the random effects (RE) regression approach to estimating volume effects. As this method does not adjust for unobserved factors that affect operative mortality, including important patient severity indicators, RE volume effect estimates may be an inappropriate measure of quality of care differences. Therefore, if our interest is in volume effects driven by quality of care differences, say for report cards, then the appropriate estimates are those from FE regression. However, if our interest is in the overall association of operative mortality with surgeon volume, either arising from quality of care or other differences, then RE volume effect estimates appear to approximate the combined FE volume effect. Note that this overall measure is the appropriate measure in assessing the impact of regionalization of CABG surgery – because the quality of care benefit from regional centers is only experienced if complicated patients would otherwise have gone to "low-volume" providers; the appropriate accounting, therefore, should combine a quality of care component with an unobserved factor component from FE regression.

Though the FE approach used here addresses some of the complexities of using observational data, a number of important limitations still remain. First, given the two levels of clustering of patients (within surgeons and hospitals), the FE approach used here is a two-step procedure, wherein the first step uses FE for one level of clustering and the second component is an RE regression at the other level. Although less restrictive than a three-tiered RE model, it nevertheless does not eliminate all forms of correlation of unobserved cluster effects. For instance, in comparing surgeons within each hospital it is assumed that patient profiles are similar across surgeons – to the extent that some of the patients are triaged to specific surgeons based on unobserved characteristics (including severity), the quality measure is confounded. Similarly, when comparing patient outcomes for the same surgeon but at different hospitals, it is assumed that there are no systematic unobserved differences among patients across hospitals.

The second limitation is that the range of clinical information to measure patients' disease burden is limited. The large unobserved mortality component estimated here might be the result of limited comorbidity information available (Charlson scores based only on ICD-9-CM codes from inpatient records for 6 months). Enriching this information, in particular, with data from clinical databases, might result in lower unobserved mortality. However, it is not clear if the volume effects associated with quality of care differences will be affected with the use of richer patient information. The FE regression used to estimate hospital volume effects uses a subset that excludes 49% of the patient records in the available data. Though we have performed a variety of sensitivity and robustness checks, there may still be other differences between those included and excluded. Third, we also recognize that this study does not attempt to disentangle the bi-directional relationship between volume and operative mortality – it only estimates a reduced form relationship between the two measures regardless of the underlying cause. Although the "learning by doing" hypothesis posits volume as the cause, an alternative hypothesis ("selective referral") allows for quality (operative mortality rate) being the cause and volume being the effect. The majority of the studies that have attempted to disentangle the causal direction, using instrumental variables regression, have found evidence for both effects at the hospital level (22-26) – we know of no studies that have modeled both surgeon and hospital volumes.

In conclusion, fixed effects (FE) regression estimates excess operative mortality from low-volume hospitals to be much higher than previously estimated. This reflects differences in operative mortality arising from quality of care differences across providers by volume. It appears that the previous estimates using random effects (RE) regression captured not only quality of care differences but also other important operative mortality determinants unobserved in the data – in particular, unobserved patient complications.

Table 1. Patient, Surgeon and Hospital Characteristics: CABG Admissions, 1998-99
(All patients are 65 or older and Medicare Fee for Service enrolled)

	All	*Within Hospital cohort**	*Within Surgeon cohort**
# CABG Admissions†	220,592	217,790	112,143
Mean operative mortality rate (per 1,000)	50.5	50.4	52.1
# Surgeons	2,772	2,744	1,216
Mean annual # CABG surgeries per surgeon	85	86	96
% Surgeons operating at 2 or more hospitals	44%	44%	100%
# Hospitals	958	898	755
Mean annual # CABG surgeries per hospital	297	311	250
% Hospitals with 2 or more surgeons	94%	100%	94%
Provider volume			
% Patients with surgeon volume <101 ("lowest tertile surgeon")‡	33%	33%	33%
% Patients with surgeon volume>162 ("highest tertile surgeon")‡	33%	33%	36%
% Patients in hospitals with volume<314 ("lowest tertile hospital")‡	33%	33%	42%
% Patients in hospitals with volume>628 ("highest tertile hospital")‡	33%	34%	28%
Patient characteristics (%)			
Age	40%	40%	40%
65 to 74	60%	60%	60%
75 to 84	37%	37%	37%
85+	3%	3%	3%
Female	35%	35%	35%
Black	4%	4%	4%
Charlson Score††	10%	10%	10%
0	41%	41%	41%
1	33%	33%	34%
2	16%	16%	16%
3+	1%	1%	1%
Nonelective admission	57%	57%	56%
Resident zip code mean Social Security income below $2,500	66%	66%	67%
Hospital characteristics (% patients with following hospital characteristics)			
Teaching	46%	46%	39%
Government owned	7%	7%	5%
Not for profit	81%	81%	80%
For Profit	11%	11%	15%

Notes:
* Within-Hospital Cohort excludes CABG admissions from 60 hospitals with only one surgeon, and Within-Surgeon Cohort retains CABG admissions from the 1,216 surgeons who operated at two or more hospitals.
† Among all these admission, there are no patients with more than one CABG surgery – each surgery admission refers to a distinct patient.
‡ Note that surgeon and hospital volumes are estimates of annual volumes covering all payers.
†† Note that the Charlson score, based on ICD-9 diagnosis codes from the index admission as well as any other inpatient admissions in the preceding 6 months, excludes primary indicators for the surgical procedure or postoperative complications.

Table 2. Operative Mortality Rate (%) for Highest & Lowest Volume Tertiles, by Sample Characteristics

	Within-Hospital Cohort (Hospitals with at least 2 surgeons)				*Within-Surgeon Cohort (Surgeons practicing at 2 or more hospitals)*			
	Hospital volume low (<314)	*Hospital volume high (>628)*	*Surgeon volume low (<101)*	*Surgeon volume high (>162)*	*Hospital volume low (<314)*	*Hospital volume high (>628)*	*Surgeon volume low (<101)*	*Surgeon volume high (>162)*
All	5.6	4.5	5.9	4.5	5.7	4.4	6.2	4.6
Age								
65-69	3.5	3.0	3.9	2.9	3.6	3.0	3.9	3.0
70-74	5.0	4.0	5.4	3.8	5.1	3.9	5.8	3.9
75-79	6.2	5.3	6.7	5.3	6.3	5.1	6.9	5.3
80-84	9.3	7.0	9.6	7.1	9.8	6.8	10.3	7.2
85+	12.1	8.5	12.9	9.0	11.5	8.2	12.7	10.3
Gender								
Female	7.0	5.6	7.4	5.4	7.2	5.6	7.9	5.5
Male	4.8	3.9	5.2	4.0	4.9	3.8	5.3	4.1
Race								
Black	6.6	5.5	6.9	5.1	6.2	4.8	6.9	4.7
Non-Black	5.5	4.5	5.9	4.4	5.7	4.4	6.2	4.6
Charlson Score								
0	5.6	4.7	6.1	4.5	5.7	4.5	6.3	4.5
1	5.2	4.1	5.4	4.2	5.3	3.9	5.8	4.3
2	5.2	4.1	5.5	4.1	5.3	4.2	5.6	4.3
3+	7.3	6.2	7.9	5.8	7.5	6.2	8.3	6.1
Resident zip code mean Social Security income								
Below $2,500	5.6	4.6	6.1	4.6	5.8	4.5	6.4	4.6
Above $2,500	5.5	4.3	5.7	4.3	5.5	4.2	5.7	4.5
Admission type								
Elective	4.2	3.4	4.4	3.3	4.5	3.4	4.7	3.5
Urgent/Emergent	6.6	5.4	7.1	5.4	6.6	5.2	7.4	5.5
Teaching status								
Teaching hospital	5.6	4.5	5.8	4.4	5.9	4.2	6.0	4.6
Not teaching hospital	5.5	4.5	6.1	4.5	5.6	4.7	6.3	4.6
Hospital ownership								
Government owned	5.6	5.3	6.0	5.1	6.2	6.5	6.5	5.4
Not for profit	5.4	4.5	5.8	4.4	5.5	4.4	6.0	4.5
For profit	6.0	3.8	6.6	4.6	6.1	3.1	6.8	4.8

Table 3. Random and Fixed Effects Regression Estimates of Excess Operative Mortality
(per 100 CABG surgeries)

	Random Effects				***Fixed Effects***			
	Within-Hospital Cohort (1)		***Within-Surgeon Cohort (2)***		***Within-Hospital Cohort (3)***		***Within-Surgeon Cohort (4)***	
	Co-eff.	***Std. Err.***	***Co-eff.***	***Std. Err.***	***Co-eff.***	***Std. Err.***	***Co-eff.***	***Std. Err.***
Hospital volume								
Lowest tertile (<314)	**0.45**	0.23	0.09	0.30			**1.36**	0.52
Middle tertile (314 - 628)	0.39	0.23	0.16	0.28			0.80	0.53
Highest tertile (>628)	Reference		Reference				Reference	
Surgeon volume								
Lowest tertile (<101)	**1.41**	0.17	**1.53**	0.26	**1.56**	0.46		
Middle tertile (101 - 162)	0.24	0.16	0.27	0.25	-0.11	0.47		
Highest tertile (>162)	Reference		Reference		Reference			
Female	**1.43**	0.10	**1.21**	0.14	**1.43**	0.10	**1.71**	0.14
Age								
65 -- 69	Reference		Reference		Reference		Reference	
70 -- 74	**1.12**	0.12	**1.06**	0.17	**1.13**	0.12	**1.22**	0.17
75 -- 79	**2.34**	0.13	**2.29**	0.18	**2.35**	0.13	**2.40**	0.18
80 -- 84	**4.51**	0.16	**4.21**	0.23	**4.51**	0.16	**4.83**	0.23
85+	**7.05**	0.30	**7.03**	0.42	**7.01**	0.30	**7.16**	0.42
Black	**0.31**	0.25	**0.77**	0.36	**0.07**	0.26	**-0.23**	0.36
Charlson score								
0	Reference		Reference		Reference		Reference	
1	**-0.52**	0.11	**-0.67**	0.15	**-0.52**	0.11	**-0.37**	0.15
2	**-0.33**	0.14	**-0.40**	0.19	**-0.31**	0.14	**-0.24**	0.20
3+	**1.75**	0.17	**1.54**	0.23	**1.77**	0.17	**2.04**	0.24
Admission type								
Elective	Reference		Reference		Reference		Reference	
Urgent/Emergent	**2.16**	0.10	**2.19**	0.14	**2.25**	0.10	**2.14**	0.14
Year								
1998	Reference		Reference		Reference		Reference	
1999	0.03	0.09	-0.03	0.13	0.08	0.10	0.06	0.14
Major teaching hospital	-0.13	0.15	-0.19	0.21			0.05	0.28
Hospital ownership								
Not for profit	Reference		Reference				Reference	
Government	0.38	0.28	0.23	0.34			0.14	0.77
For profit	0.61	0.24	0.66	0.41			-0.16	0.45
Resident zip code mean Social Security income below $2,500	0.16	0.11	0.18	0.15	0.05	0.11	0.05	0.16

Note: Estimates in bold are significant at 5% level.

Table 4. Expected Operative Mortality Rates (%) by Hospital & Surgeon Volume

	(1) Quality of care component[†] (FE Expected)	(2) Unobserved Factor component (FE Residual)	(3) Total FE (1) + (2)	(4) RE[†]
Hospital volume				
Lowest tertile (<314)	1.36	-0.84	0.52	0.45
Middle tertile (314 - 628)	0.80	-0.40	0.40	0.39
Highest tertile (>628)	Reference	Reference	Reference	Reference
Surgeon volume				
Lowest tertile (<101)	1.56	-0.58	0.98	1.41
Middle tertile (101 - 162)	-0.11	0.07	-0.04	0.24
Highest tertile (>162)	Reference	Reference	Reference	Reference

† These figures are from Table 3 (FE expected values are from columns 3 or 4, whereas RE values are from column 1).

REFERENCES

1. Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon volume and operative mortality in the United States. N Engl J Med 2003;349:2117-2127
2. Shahian DM, Normand S-LT. The volume-outcome relationship: from Luft to Leapfrog. Ann Thorac Surg 2003;75:1048-1058
3. Halm EA, Lee C, Chassin MR. Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature. Ann Intern Med 2002;137:511-520
4. Luft HS, Bunker J, Enthoven A. Should operations be regionalized? The empirical relation between surgical volume and mortality. N Engl J Med 1979;301:1364-1369
5. Halm EA, Lee C, Chassin MR. How is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature. Washington, D.C.: Institute of Medicine; 2000
6. Hewitt M. Interpreting the Volume-Outcome Relationship in the Context of Health Care Quality: Workshop Summary. Washington, D.C.: Institute of Medicine; 2000
7. Birkmeyer JD. Should we regionalize major surgery? potential benefits and policy considerations. Journal of the American College of Surgeons 2000;190:341-349
8. Birkmeyer JD. Understanding Surgeon Performance and Improving Patient Outcomes. J Clin Oncol 2004;22:2765-2766
9. Fredenheim M. To Find a Doctor, Mine the Data. The New York Times. New York City; 2005
10. Birkmeyer NJO, Birkmeyer JD. Strategies for Improving Surgical Quality -- Should Payers Reward Excellence or Effort? N Engl J Med 2006;354:864-870
11. Dranove D, Kessler D, McClellan M, et al. Is more information better? The effects of "report cards" on health care providers. Journal of Politcal Economy 2003;111:555
12. Jha AK, Orav EJ, Li Z, et al. Concentration and Quality of Hospitals That Care for Elderly Black Patients. Arch Intern Med 2007;167:1177-1182
13. Losina E, Wright EA, Kessler CL, et al. Neighborhoods matter: use of hospitals with worse outcomes following total knee replacement by patients from vulnerable populations. Arch Intern Med 2007;167:182-187
14. Gowrisankaran G, Town RJ. Estimating the quality of care in hospitals using instrumental variables. Journal of Health Economics 1999;18:747-767
15. Hamilton BH, Ho V. Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. Medical Care 1998;36:892-903
16. Birkmeyer JD, Siewers AE, Finlayson EVA, et al. Hospital Volume and Surgical Mortality in the United States. N Engl J Med 2002;346:1128-1137
17. Raudenbush SW, Bryk AS. Hierarchical Linear Models. Thousand Oaks, CA: Sage Publications; 2002
18. StataCorp. Stata Statistical Software: Release 9. College Stataion, TX: StataCorp LP; 2005
19. Johnston J, DiNardo J. Econometric Methods. McGraw-Hill; 1997

20. Schneider EC, Epstein AM. Influence of Cardiac-Surgery Performance Reports on Referral Practices and Access to Care -- A Survey of Cardiovascular Specialists. N Engl J Med 1996;335:251-256
21. Schwartz LM, Woloshin S, Birkmeyer JD. How do elderly patients decide where to go for major surgery? Telephone interview survey. BMJ 2005;331:821-
22. Gowrisankaran G, Ho V, Town R. Causality, Learning and Forgetting in Surgery. Working Paper, John M Olin School of Business, Washington University in St Louis. St. Louis; 2005
23. Tsai AC, Votruba M, Bridges JFP, et al. Overcoming Bias in Estimating the Volume-Outcome Relationship. Health Services Research 2006;41:252-264
24. Farley DE, Ozminkowski RJ. Volume-outcome relationships and in-hospital mortality: the effect of changes in volume over time. Med Care 1992;30:77-94.
25. Luft HS. The relation between surgical volume and mortality: an exploration of causal factors and alternative models. Med Care 1980;18:940-959.
26. Gaynor M, Seider H, Vogt WB. The volume-outcome effect, scale economies, and learning-by-doing. American Economic Review 2005;95

PUBLICATIONS AND PRODUCTS

Hanchate A, Stukel T, Birkmeyer J and Ash A. Surgery volume, quality of care and operative mortality in coronary artery bypass graft surgery: A re-examination using fixed-effects regression. (submitted to *Health Services and Outcomes Research Methodology*). Oct 2007

LECTURES AND PRESENTATIONS

"Surgery Volume, Quality of Care and Operative Mortality" AcademyHealth Annual Research Meeting, Orlando, FL, June 3-5, 2007

"Surgery Volume and Mortality: A Re-examination Using Fixed-Effects Regression." AcademyHealth Annual Research Meeting, Seattle, WA, June 26, 2006 (Selected as *2006 Most Outstanding Abstracts*, AcademyHealth 2006)

"Surgery Volume and Mortality: Re-estimation using fixed-effects methods," 2006 New England SGIM Regional Meeting, Boston, MA, March 24, 2006