

Measuring the Volume-Outcome Relation for Complex Hospital Surgery

Woohyeon Kim¹ · Stephen Wolff² · Vivian Ho^{1,3,4}

Published online: 15 April 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Background Prominent studies continue to measure the hospital volume-outcome relation using simple logistic or random-effects models. These regression models may not appropriately account for unobserved differences across hospitals (such as differences in organizational effectiveness) which could be mistaken for a volume outcome relation.

Objective To explore alternative estimation methods for measuring the volume-outcome relation for six major cancer operations, and to determine which estimation method is most appropriate.

Methods We analyzed patient-level hospital discharge data from three USA states and data from the American Hospital Association Annual Survey of Hospitals from 2000 to 2011. We studied six major cancer operations using three regression frameworks (logistic, fixed-effects, and random-effects) to determine the correlation between patient outcome (mortality) and hospital volume.

Results For our data, logistic and random-effects models suggest a non-zero volume effect, whereas fixed-effects models do not. Model-specification tests support the fixed-effects or random-effects model, depending on the surgical procedure; the basic logistic model is always rejected. Esophagectomy and rectal resection do not exhibit significant volume effects, whereas colectomy, pancreatic resection, pneumonectomy, and pulmonary lobectomy do.

Conclusions The statistical significance of the hospital volume-outcome relation depends critically on the regression model. A simple logistic model cannot control for unobserved differences across hospitals that may be mistaken for a volume effect. Even when one applies panel-data methods, one must carefully choose between fixed- and random-effects models.

Electronic supplementary material The online version of this article (doi:[10.1007/s40258-016-0241-6](https://doi.org/10.1007/s40258-016-0241-6)) contains supplementary material, which is available to authorized users.

✉ Vivian Ho
vho@rice.edu

¹ Department of Economics MS-22, Rice University,
6100 Main Street, Houston, TX 77005, USA

² Department of Mathematics MS-136, Rice University,
6100 Main Street, Houston, TX 77005, USA

³ Baker Institute for Public Policy, Rice University,
6100 Main Street MS 40, Houston, TX 77005, USA

⁴ Department of Medicine, Baylor College of Medicine,
Houston, TX, USA

Key Points for Decision Makers

We illustrate (i) how to apply the fixed-effects and random-effects regression frameworks and (ii) how to determine which regression framework is most appropriate for given data.

We find that both random-effects and fixed-effects model are more appropriate than a simple logistic model for measuring the volume-outcome effect. For four operations, the random-effects model is sufficient. However, for two operations, the fixed-effects model is more appropriate.

Policy makers who may be considering the centralization of complex operations to improve patient outcomes may falsely conclude that a volume-outcome relation exists, if decisions are based on analysis from simple logistic models.

Implementation of panel-data methods (like the fixed-effects and random-effects frameworks) following the example in this paper may lead to more reliable policy recommendations.

1 Introduction

For years, numerous studies have asserted a positive correlation between hospital or surgeon volume and patient outcome. Halm et al. [1] systematically reviewed over 130 volume-outcome studies, approximately 70 % of which found a significant volume effect. Halm et al.'s review was published in the *Annals of Internal Medicine*, and the majority of volume-outcome studies have also been published in clinical journals. Based on findings such as these, many researchers and policy groups advocate centralizing procedures at a small number of hospitals, in order to take advantage of the volume effect [2].

The majority of volume-outcome studies to date use simple regression models, such as basic logistic regressions, that ignore omitted-variable bias. This approach potentially leads to spurious conclusions and improper policy recommendations. For example, suppose that certain large-volume hospitals are particularly effective at organizing surgical teams that reduce complications. A regression that does not control for organizational skill will find a positive association between hospital volume and patient outcome. However, it is organizational expertise, not higher hospital volume, that drives improved patient outcomes.

Panel data sets provide means to help control for omitted-variable bias, specifically via the fixed-effects and

random-effects¹ regression frameworks. Both frameworks control for time-invariant heterogeneity, i.e. unobserved fixed factors that differ across hospitals. The random-effects framework does this under the assumptions that this heterogeneity is uncorrelated with the other explanatory variables and follows a known distribution. The fixed-effects framework is more general, allowing correlation with other explanatory variables and making no parametric assumption about the distribution.

The tendency to use basic logistic models when estimating the volume effect continues to this day. Searching Google Scholar and PubMed with keywords “volume-outcome relationship” for publications from 2008 to 2013, we located 87 peer-reviewed studies. Only 11 employed panel-data methods, and only three used fixed-effects models. In addition, studies using fixed-effects and random-effects models often fail to validate that the assumptions underlying these frameworks are satisfied by the data. These publication trends are not limited to smaller journals: for example, the *New England Journal of Medicine* in 2010, 2011, and 2013 published studies on the volume effect that do not explicitly consider fixed-effects models [3–5].

Cancer procedures are of particular interest regarding improved outcome and efficiency. According to the American Cancer Society, cancer is the second-leading cause of death in the USA, “accounting for nearly one of every four deaths” [6]. In addition to the human toll, the National Institutes of Health estimates that cancer cost the US economy approximately US\$86.6 billion in direct medical costs in 2009 and US\$130 billion in lost productivity due to premature death [6].

Previous research on the volume-outcome relationship in cancer procedures is conflicted. The survey by Halm et al. [1] mentions that nine of ten studies on pancreatic cancer and three of three studies on esophageal cancer find a significant volume effect at the hospital or surgeon level. Several recent clinical studies, dating from 2008 to 2012, report similar findings [4, 7–10]. However, Lapar et al. [11] finds no significant association between hospital procedure volume and patient mortality for several surgical procedures, including pancreatic and esophageal resection.

This study addresses three major questions. First, does a volume effect exist in any of six major cancer resection procedures? Second, how does the regression framework used affect this answer? Third, for a given data set, which regression framework is most appropriate? To address these questions, we analyze a 12-year panel data set of hospital-discharge data of patients who underwent one of six cancer procedures. We fit the data for each of these

¹ In this paper, “random-effects model” means a model whose intercept includes a random effect.

procedures to basic logistic, fixed-effects logistic, and random-effects logistic regressions. We could have taken a different approach, using Monte Carlo simulation to generate samples with and without volume–outcome effects and testing whether each of the three estimation approaches correctly identify the presence or absence of a relationship between procedure volume and outcomes. We chose instead to focus on an application involving actual clinical data. This approach is more relevant to clinicians and policy makers, who are most likely to shape future decisions on whether or not to centralize complex care.

2 Previous Literature

The vast majority of studies dealing with binary patient outcomes such as mortality employ a simple logistic regression framework. A few studies have used the random-effects model [3, 12–16]. However, most studies using the random-effects model do not check whether their data satisfy the assumptions of the random-effects model, nor do they test alternative model specifications. Moreover, most do not discuss omitted-variable bias as justification for their model choice.

The fixed-effects regression framework uses the variation within a group to exclude omitted-variable bias from time-invariant factors [17]. With a few exceptions [18–20], the fixed-effects model is rarely used in the volume–outcome literature. However, previous research confirms that controlling for unobserved heterogeneity by using a fixed-effect model can yield drastically different results. In a study of child immunization in China, Xie et al. [21] uses a fixed-effects model to control for community- or household-level time-invariant characteristics, rendering insignificant what appeared to be a significant effect of wealth. In a study of hip-fracture patients, Hamilton et al. [22] also find that a significant volume effect disappears when using a fixed-effects model.

It is crucial for researchers to correctly measure volume–outcome effects, because there are potential unintended consequences of centralization. Kessler et al. [23] and others have shown that hospital competition tends to enhance patient welfare. Centralization typically reduces the competitiveness of healthcare markets. It is not clear a priori whether the benefit (if any) from a volume effect would outweigh the welfare loss associated with reduced competition.

In his seminal 1979 paper identifying an empirical relation between surgical volume and mortality for 12 different operations, Luft et al. [24] cautioned that the observed volume–outcome association could be attributable to selective referral; more patients may be drawn to hospitals that have better outcomes. More

recently, economists have applied instrumental variables analysis to distinguish between volume driving patient outcomes (a learning by doing effect), versus better outcomes leading to higher volume (selective referral). In these studies, distance to providers or the number of patients and other hospitals within close vicinity of a particular hospital are used as instruments for hospital volume that are unlikely to be confounded by selective referral [25, 26]. These studies find that the proposed instruments are valid predictors of hospital volume. Hypothesis tests also reveal no evidence for selective referral. Another paper conducts hypothesis tests for the exogeneity of hospital volume in explaining patient mortality and finds no evidence that the volume–outcome relation is the result of selective referral [27].

Given that past studies that test for patient selection in the volume–outcome relation find no evidence of selective referral, we chose not to apply instrumental variables analysis in this paper. Instead, we focus on comparing random- and fixed-effects models. The majority of clinicians consult clinical journals when they seek to learn whether a volume–outcome relation exists for a particular operation [1, 28]. And the overwhelming majority of clinical studies apply a simple logistic regression to test for a volume–outcome relation, while most of the rest apply random-effects analysis. Estimation with fixed effects can be readily applied to the same data sets that have been analyzed in these published studies. It is critical for clinicians and policy makers to know whether failing to control for potentially systematic but unobservable differences between high- and low-volume hospitals can yield misleading conclusions regarding the presence of a volume–outcome effect.

3 Methods

3.1 Study Population

We use hospital-discharge data from Florida, New Jersey, and New York for the 12 years 2000–2011, containing all patients discharged during this time.² Based on previous literature [29], we selected patients with an International Classification of Diseases 9th Revision, Clinical Modification (ICD-9-CM) cancer diagnosis and a principal procedure code for colectomy, esophagectomy, pancreatic resection, pneumonectomy, pulmonary lobectomy, or rectal resection. We apply three additional inclusion criteria: the patient was 21 years or older at admission, the patient was not transferred to another hospital (as opposed to being

² All data were obtained from the respective states' Departments of Health.

discharged home or to a rehabilitation hospital, or died during the hospital stay), and the patient was successfully matched with hospital-level data from the American Hospital Association (AHA) Surveys.³ Because the state-level discharge data do not provide a persistent patient-specific identifier, we could not control for multiple admissions of the same patient within a year. However, analysis suggests that this problem, if present, is negligible.⁴ Annual hospital-level data were obtained from the AHA Surveys, 2000–2011.

3.2 Variables

Our outcome measure is in-hospital mortality. To define hospital volume, we compute the total number of patients treated by each hospital for each procedure within each year. Volumes are computed before applying inclusion criteria to avoid endogeneity.⁵ Hospital-specific characteristics for each year include total facility expenses, full-time-equivalent physicians, ownership status, teaching status, and urban status. Patient characteristics include admission status, age, cancer stage, Elixhauser co-morbidities, race, and sex.

3.3 Statistical Analysis

For each procedure, we fit three models with cluster-robust standard errors: logistic, conditional fixed-effects logistic,⁶ and random-effects logistic. For each model, the log odds of mortality is regressed on patient characteristics, hospital characteristics, and indicator variables for each subprocedure (if any) and year.

In fitting the fixed-effects model, estimating an indicator variable for each hospital would lead to inconsistent estimates, known as the incidental-parameters problem. To avoid this, we use the conditional logistic distribution suggested by Chamberlain [30]. By conditioning the likelihood function on the sum of the dependent variables, a sufficient statistic, we obtain a conditional likelihood function that does not depend on the hospital indicator

³ Patients under the age of 21 years are dropped to avoid issues arising from different patterns of cancer treatment for children and adults. Transfer-out patients are dropped to avoid endogeneity issues. For each procedure, less than 0.5 % of patients were unlinked to hospital-level data.

⁴ Specifically, for each procedure and within each year, we count the number of observations whose age, sex, and state agreed (an estimate for an upper bound on the number of multiple admissions). The maximum number of duplicate patients is at most 0.10 % of a procedure's sample; see Appendix 2 in the Supplemental Material.

⁵ Surgeon volume, used in some robustness-check specifications, is defined analogously.

⁶ The conditional likelihood function, introduced in [30], avoids the incidental-parameter problem.

variables.⁷ Thus, to fit the fixed-effects model, we do not have to estimate the hospital indicator variables, allowing us to avoid the incidental-parameters problem. In Stata, this approach is effected by the `-clogit-` command.

Two aspects of our analysis merit brief discussion. First, we do not include surgeon fixed effects. The cancer procedures analyzed here typically have very low surgeon volume: the median surgeon volume, summed over the 12-year sample period, ranged from one (esophagectomy) to seven (lobectomy). As one might expect with such low volumes, the percentage of surgeons having nontrivial variation in outcome (i.e., at least one patient survives and does not survive the operation during the sample period) does not exceed 15 % in any of the surgical procedures; these surgeons account for no more than 25 % of observations in the respective procedure. Because the fixed-effects model uses only within-group variation, including surgeon fixed effects would have required discarding more than 75 % of the data.⁸

Second, we run each model specification both with and without a surgeon-volume variable.⁹ Most existing studies of the hospital-level volume effect omit surgeon volume, so to facilitate comparison, we focus on results from the regressions without surgeon volume. This specification risks misattributing a surgeon-level volume effect to the hospital level. For our data, the main findings are robust under both specifications.

3.4 Model Specification Tests

After fitting the three models, we perform model-specification tests. To compare the fixed-effects and random-effects models, one typically uses the Hausman test. However, the standard Hausman test is invalid for clustered data [33]: the clustered structure causes the error terms of observations in the same group not to be independently distributed, which in turn implies that the random-effects estimator is not efficient. Several alternative approaches have been proposed; we adopt the approach proposed by Mundlak [34]. We briefly present this approach in the linear setting, for simplicity. (The extension to the

⁷ See section 17.4.4 in Greene [31].

⁸ If surgeons perform all or nearly all of their procedures in a single hospital, then including both hospital and surgeon fixed effects in a regression will cause problems of multicollinearity. In this case, if identifying both effects is important, one can use a two-way nested model: e.g., a fixed term for the hospital effect and a random term for the surgeon effect. See Schielzeth et al. [32].

⁹ Because surgeons can perform procedures across multiple hospitals, an ideal setting would account for correlation of error terms at both the hospital and physician level. Stata does not currently appear to handle clustered standard errors under multi-level clustering. Due to this limitation, in our models we implement cluster-robust standard errors at the hospital level only.

nonlinear case is similar) [31]. In our study, we apply this approach within the logistic framework.

Consider nested panel data (patients, indexed by i , are clustered within hospitals, indexed by j , which are observed over time t), and the basic linear regression model $Y = X\beta + \varepsilon$.

We assume that the error term ε_{ijt} can be decomposed into

$$\varepsilon_{ijt} = c_j + d_t + u_{ijt}$$

where c_j is a constant capturing time-invariant unobserved characteristics of hospital j , d_t is a time fixed effect, and u_{ijt} is the error term associated with patient i in hospital j at time t . The random-effects model assumes that c_j is uncorrelated with the other regressors X_{ijt} . This assumption allows us to treat c_j as a random variable conditional on X_{ijt} . If the unobserved characteristics c_j are correlated with the other regressors, then the estimates of the random-effects model are inconsistent. The fixed-effects model makes no assumption about the correlation between c_j and the other regressors; in particular, they are allowed to be correlated.

Mundlak [34] proposes the following approach. Let Z be the subset of hospital-specific variables in X . For each hospital j , we take the time average of the observed hospital characteristics, $\bar{Z} = \frac{1}{T} \sum_{t=1}^T Z_{jt}$, and include these averages in the random-effects model, obtaining

$$Y = X\beta + \bar{Z}\gamma + \varepsilon. \tag{1}$$

Under the assumption that

$$E[c_j|Z_j] = \bar{Z}_j\gamma \tag{2}$$

$\gamma = 0$ implies that c_j and Z_j are uncorrelated.¹⁰ Empirically, then, fitting model (1) and performing a significance test on γ gives us a way to compare the fixed-effects and random-effects models: A γ significantly different from zero implies that the noncorrelation assumption underlying the random-effects model fails to hold. In this case, the fixed-effects model is preferred.

We compare the fixed-effects and pooled logistic models by fitting a seemingly unrelated regression. Let b and \tilde{b} denote the coefficients of the fixed-effects and pooled logistic models, respectively. We estimate $\text{Cov}(b, \tilde{b})$ via a seemingly unrelated regression assuming correlation of the error terms in the two models, then use this result to compare $\text{Var}(b - \tilde{b})$ via its expansion $\text{Var}(b) + \text{Var}(\tilde{b}) - 2\text{Cov}(b, \tilde{b})$. This variance is used to construct the usual Wald-type test statistic for comparing two models.

¹⁰ Given $\gamma = 0$ and (2), the law of iterated expectations and the definition of correlation yield the stated result.

4 Results

4.1 Descriptive Statistics

After applying the three inclusion criteria in section 3.1, we obtained 164,804 cancer patients hospitalized for colectomy, 4827 for esophagectomy, 14,246 for pancreatic resection, 5043 for pneumonectomy, 54,448 for pulmonary lobectomy, and 36,046 for rectal resection.¹¹ Table 1 provides descriptive statistics for these data.

The average patient age is between 63 and 71 years. The racial composition is 76–86 % White, 4–11 % Black, 5–9 % Hispanic, and 5–9 % other. Females comprise roughly half of the patient population for all procedures except esophagectomy (19.35 % female) and pneumonectomy (36.51 % female). The percentage of patients with nodal cancer is 16–40 %; the percentage of patients with metastasized cancer is 7–28 %. Counting hospitals in different years as unique (to allow for changes in hospital status), public ownership ranges from 10 to 15 %, and teaching status ranges from 16 to 42 %.

The fixed-effects framework estimates the volume–outcome relationship using only within-hospital variation, ignoring variation across hospitals. If most hospitals experience little volume variation over the sample period, one will obtain noisy estimates. Figures 1 and 2 present measures of within-hospital variation in volume for our data. The coefficients of variation plotted in Fig. 1 show the relative size of standard deviation and mean volume within each hospital. In general, these coefficients of variation are around 0.5; that is, a given hospital tends to have a standard deviation equal to about half its mean. The histograms of absolute deviation from mean hospital volume (computed separately for each hospital) in Fig. 2 show that for all procedures, our data have relatively good coverage of deviations between 0 and the sample mean (i.e., the mean volume over all hospitals). Moreover, for these procedures, we observe a reasonable number of deviations greater than the sample mean. Using Stata’s `-xtsum-` command, we find that the within-hospital variation is at least 29 % for each of the six procedures.¹² These results suggest that our data possess sufficient within-hospital variation to allow us to credibly fit the fixed-effects model.

¹¹ Before the three inclusion criteria in 3.1 were applied, the number of patients for each procedure was as follows. Colectomy: 165,729; esophagectomy: 4854; pancreatic resection: 14,357; pneumonectomy: 5065; pulmonary lobectomy: 54,620; rectal resection: 36,199.

¹² The exact within-hospital variations for each procedure are as follows: Colectomy: 30.11 %; esophagectomy: 37.92 %; pancreatic resection: 33.88 %; pneumonectomy: 43.75 %; pulmonary lobectomy: 29.60 %; rectal resection: 37.39 %.

Table 1 Descriptive statistics

Variable	Colectomy	Esophagectomy	Pancreatic resection	Pneumonectomy	Pulmonary lobectomy	Rectal resection
Number of patients	164,804	4827	14,246	5043	54,448	36,046
Length of stay (mean), days	10.21	16.74	15.67	8.92	8.41	9.31
Mortality (mean), %						
2000						
Small	8.20	13.89	14.94	15.23	7.00	0
Medium	3.67	10.71	10.10	5.88	4.22	1.98
Large	2.50	5.56	4.44	10.20	2.33	0.93
2011						
Small	2.48	9.38	2.27	5.88	4.44	2.52
Medium	2.37	14.29	8.33	2.56	3.15	2.14
Large	1.96	0	5.41	17.39	0	1.19
Number of hospitals	520	235	377	343	425	500
Publicly owned, %	10.66	14.06	14.27	11.33	10.06	10.29
Teaching hospital, %	16.89	41.11	30.23	27.16	20.43	18.25
Volume (hospital)						
2000–2011						
Mean ^a	27.94	2.60	3.93	2.02	11.51	6.67
Std dev ^a	7.02	1.41	1.97	0.94	4.22	2.86
2000						
Mean	33.10	3.48	4.28	2.87	12.58	9.06
Max	198	56	126	42	304	151
2011						
Mean	27.76	6.00	9.44	2.67	15.46	6.54
Max	238	87	149	17	263	122
Age (mean), years	70.65	63.41	65.77	63.31	67.61	66.20
Race						
Black, %	10.79	4.58	9.09	5.97	6.13	7.52
Hispanic, %	8.11	6.86	8.67	5.61	5.02	8.41
White, %	76.88	85.91	76.82	83.98	85.00	78.74
Other, %	6.84	5.84	8.73	6.96	5.87	8.17
Sex, female, %	52.76	19.35	49.33	36.51	51.83	43.09
Stage of cancer						
Nodal, %	27.35	29.58	36.21	39.90	16.90	25.33
Metastasized, %	18.22	8.89	27.54	15.45	7.85	14.38

The mortality measure used is in-hospital mortality, i.e., a discharge code of 20 in the hospital-discharge data. For this table, we define “small,” “medium,” and “large” hospitals using quartiles within the given year: small hospitals are defined as those that fall in the lowest quartile; medium hospitals, in the middle two quartiles, and large hospitals, in the highest quartile. “Number of patients” excludes patients less than 21 years of age at the time of admission, transfer-out patients, and patients unmatched to hospital-level data

^a The mean and standard deviation presented here are computed by first computing the mean and standard deviation of volume within each hospital, then taking the mean of these values over all hospitals in the sample (with each hospital receiving equal weight)

4.2 Regression Models and Analysis

Coefficient estimates for the explanatory variables in all three regressions (basic logistic, fixed-effects, and random-effects) for each procedure are in the Supplementary Material. Neither teaching status nor public ownership

appear to significantly affect in-hospital mortality. All regression frameworks find the probability of mortality increases with age. The regressions reveal a small number of significant correlations between race and mortality, but no clear trend emerges either within or among procedures. Cancer stage is significantly correlated with an increased

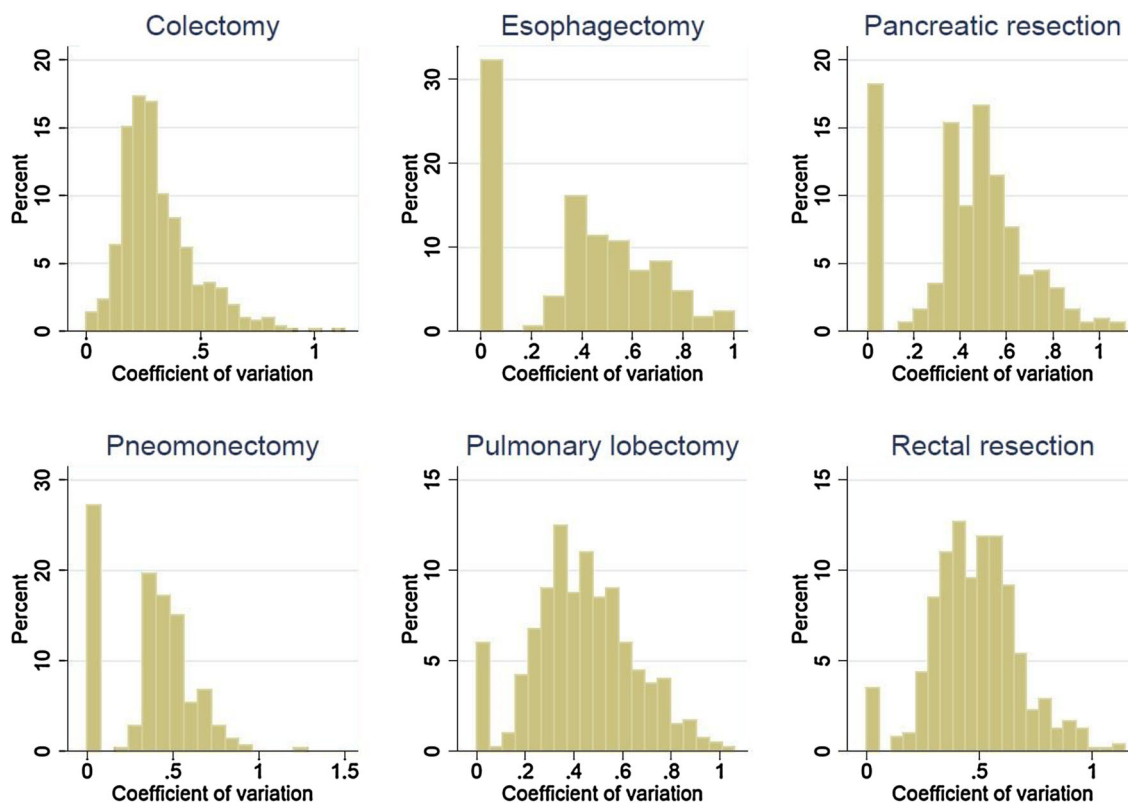


Fig. 1 Coefficient of variation for hospital volume (within-hospital). For each hospital, the coefficient of variation is computed by dividing the standard deviation of volume for the hospital by the hospital’s

mean. The values are plotted here as histograms. See Appendix 4 in the Supplemental Material for quantiles of hospital volume for each procedure

probability of mortality for only half of the procedures studied here (metastatic: colectomy, pulmonary lobectomy, rectal resection).

Of particular interest are the coefficients on hospital volume and their discrepancies among the regression frameworks. The estimates for the hospital volume variable are presented in Table 2. For all procedures except esophagectomy (for which all regression models find no significant volume effect), the logistic and random-effects regressions find a volume effect with $p < 0.05$ or better. The direction of this effect is consistent with previous findings in the literature: higher volume is correlated with lower mortality. In contrast, the fixed-effects regressions find no significant volume effect in any of the procedures.

Given the differences between models, we implement the two specification tests described above. In Table 3, the first test compares the fixed-effects and random-effects models: we run the regression proposed by Mundlak [34], followed by a test of the null hypothesis that all mean variables of hospital characteristics are jointly zero. In this case, rejection of the null hypothesis favors the fixed-effects model. For two of the six procedures—esophagectomy and rectal resection—we reject the null hypothesis with $p < 0.0132$ and $p < 0.0163$, respectively. The second

test assesses unobserved heterogeneity by comparing the logistic and fixed-effects models: we first run a seemingly unrelated regression using the two models, implemented via Stata’s `-suest-` command, followed by a test of the null hypothesis that the coefficients of all variables common to both models are identical. For all six procedures, this test strongly rejects the null hypothesis: for colectomy, $p < 0.0037$, and for the other five procedures, $p < 0.0001$. These results imply that the hospital-specific indicator variables introduced by the fixed-effects framework capture relevant unobserved heterogeneity.

As shown in Table 2, the standard errors on the hospital-volume variable in the fixed-effects regressions are typically within one to two times the magnitude of those in the random-effects regressions. Note also that the point estimates of the volume coefficient move closer to zero (typically twice as close to zero) in the fixed-effects regression compared to the random-effects regression. This suggests that the absence of significant volume effects in the fixed-effects models is not due to insufficient within-group variation.

We conclude that, of the three regression frameworks analyzed here, the fixed-effects framework is most appropriate for esophagectomy and rectal resection, while the

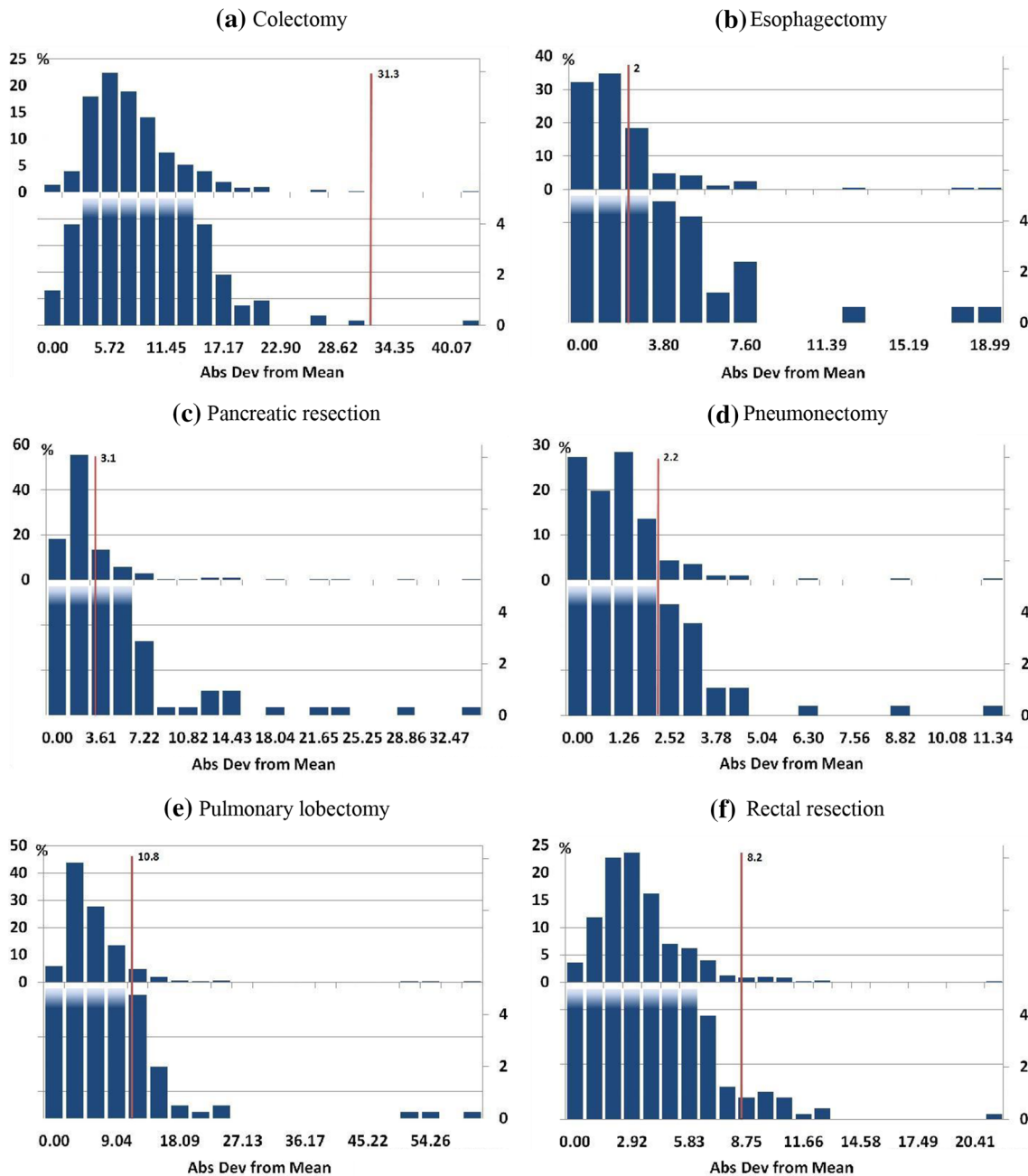


Fig. 2 Absolute deviation from mean of hospital volume (within-hospital). For each hospital, for each year the hospital is active (i.e., performs at least one surgery) we compute the absolute value of the difference between the hospital’s yearly volume and its mean over the 12-year period. We plot all results here as histograms. The mean

hospital mean volume, aggregated over all hospitals (counting each hospital once), is plotted in red. For each procedure, the bottom plot zooms in on small values of the vertical axis, offering a better view of the right tail

random-effects framework is most appropriate for the other four procedures. In particular, this implies that we find no significant volume effect for esophagectomy and rectal resection. These two procedures have comparable sample sizes to other surgical procedures studied here, so this result does not seem to be a product of insufficient data. Similarly, neither the in-hospital mortality rate nor the

magnitude of the volume coefficient in these two procedures differs drastically from those in the other procedures.

4.3 Robustness Checks

We subject our findings to several robustness checks. To investigate the possibility that the volume effect is

Table 2 Coefficient and standard error on hospital volume

Procedure	Regression model			<i>N</i>	
	Logistic regression	Random-effects model	Fixed-effects model	Logistic regression, Random-effects model	Fixed-effects model
Colectomy	−0.0041*** (0.0009)	−0.0038*** (0.0010)	−0.0025 (0.0016)	164,204	163,447
Esophagectomy	−0.0039 (0.0068)	−0.0076 (0.0089)	0.0010 (0.0086)	4785	4239
Pancreatic resection	−0.0116*** (0.0030)	−0.0118*** (0.0025)	−0.0054 (0.0044)	14,166	13,469
Pneumonectomy	−0.0248** (0.0080)	−0.0237* (0.0093)	−0.0093 (0.0194)	5016	4087
Pulmonary lobectomy	−0.0031*** (0.0008)	−0.0036*** (0.0011)	−0.0020 (0.0017)	54,351	51,239
Rectal resection	−0.0072** (0.0026)	−0.0070* (0.0027)	−0.0039 (0.0091)	35,833	28,932

These regressions results are from the sample excluding patients less than 21 years of age at the time of admission, transfer-out patients, and patients unmatched to hospital-level data

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3 Model specification tests

	Colectomy	Esophagectomy	Pancreatic resection	Pneumonectomy	Pulmonary lobectomy	Rectal resection
-suest-						
<i>df</i>	61	59	61	57	59	62
χ^2	94.64	3693.65	1583.33	270,000.00	1184.64	4769.47
<i>p</i> value	0.0037	0.0000	0.0000	0.0000	0.0000	0.0000
Mundlak						
<i>df</i>	7	7	7	7	7	7
χ^2	3.08	17.74	4.52	6.51	13.36	17.18
<i>p</i> value	0.8772	0.0132	0.7185	0.4817	0.0638	0.0163

To compare the logit versus fixed-effects specifications, we use Stata's `-suest-` command to fit a seemingly unrelated regression model using the logit and fixed-effects models; we then use the `-test-` command to test the equality of all coefficients common to the two regressions. To compare the fixed-effects versus random-effects specifications, we fit a Mundlak model (described in Sect. 3.3); we then use the `-test-` command to test whether the four hospital variables are jointly zero in this model

nonlinear, we fit the regression models using linear splines with four knots (see Appendix 1 in the Supplemental Material). In almost all cases, no statistically significant difference is found between slopes in adjacent regions; in addition, in all cases, the coefficient of the highest-volume region is not significantly different from the coefficient of the lowest-volume region. These results indicate no concavity, suggesting that our linear specification of volume is reasonable.

Despite this absence of concavity, as an additional check we fit the regression models using the square root of hospital volume, following Seider et al.¹³ The results correspond closely to those of the main analysis in Table 2. In particular, for rectal resection, model specification tests continue to favor the fixed-effects framework, which does not find a volume effect, over the random-effects framework, which does. We also fit the regression models including a surgeon-level volume variable. Again we

observe discordant findings for volume among the specifications, and again the fixed-effects regression is preferred in the case of rectal resection, with no significant volume effect. All of these results are presented in Table 4.

5 Discussion

Many studies have measured the relationship between hospital volume and outcome. However, different studies report contradictory findings, and few studies report results from more than one model specification. This study uses a 12-year panel data set to investigate the volume–outcome relationship for six cancer procedures, applying three different regression frameworks, and tests their validity.

For our data, the basic logistic regression model—the most commonly used model in the literature to date—is always strongly rejected in favor of models that control for time-invariant heterogeneity. A study of coronary artery bypass grafts by Huesch [25] finds similar results, always rejecting the basic logistic model. For our data the random-

¹³ See the working paper, Seider, Gaynor, Vogt. Volume–outcome and Antitrust in US Health Care Market. 2004.

Table 4 Robustness checks

Procedure	The square root of hospital volume			Including surgeon volume			N
	Logistic regression	Random-effects model	Fixed-effects model	Logistic regression	Random-effects model	Fixed-effects model	
Colectomy	-0.0558*** (0.0128)	-0.0510*** (0.0127)	-0.0339 (0.0239)	-0.0040*** (0.0009)	-0.0038*** (0.0009)	-0.0024 (0.0016)	164,204
Esophagectomy	-0.1108 (0.0703)	-0.1502 (0.0788)	-0.0381 (0.1169)	0.0034 (0.0045)	0.0030 (0.0081)	-0.0029 (0.0072)	4785
Pancreatic resection	-0.1820*** (0.0358)	-0.1887*** (0.0327)	-0.0854 (0.0627)	-0.0073*** (0.0021)	-0.0085*** (0.0020)	-0.0026 (0.0045)	14,166
Pneumonectomy	-0.1370* (0.0580)	-0.1223 (0.0696)	0.0023 (0.1200)	-0.0173 (0.0136)	-0.0155 (0.0154)	-0.0011 (0.0255)	5016
Pulmonary lobectomy	-0.0681*** (0.0148)	-0.0725*** (0.0173)	-0.0358 (0.0373)	-0.0016* (0.0007)	-0.0017 (0.0009)	0.0004 (0.0019)	54,351
Rectal resection	-0.0730* (0.0336)	-0.0692* (0.0352)	-0.0219 (0.0770)	-0.0065* (0.0028)	-0.0064* (0.0029)	-0.0036 (0.0091)	35,833

For robustness check, (1) we use the square root of hospital volume as our main independent variable instead of hospital volume, and (2) include surgeon volume along with hospital volume. These regressions results are from the sample excluding patients less than 21 years of age at the time of admission, transfer-out patients, and patients unmatched to hospital-level data

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

effects model always finds a significant volume effect (except for esophagectomy), whereas the fixed-effects model never does. Whether the fixed-effects or random-effects model is preferred varies across surgical procedures.

We find no significant volume effect for esophagectomy and rectal resection; we find a precisely estimated volume effect for the other four cancer operations. Esophagectomy is the least common procedure in our sample ($N = 4827$), and rectal resection is the third most common ($N = 36,046$). Esophagectomy has some of the highest mortality rates in the sample (comparable to pancreatic resection), whereas mortality rates for rectal resection are the lowest. The difference in nature between these two operations suggests that both infrequent and common operations, as well as operations of differing difficulty, should be carefully examined for a volume effect.

While not the focus of this paper, for those procedures that display a significant volume effect, one may ask whether this association results from a practice-makes-perfect mechanism (also known as learning by doing) or a selective-referral mechanism. Under the practice-makes-perfect hypothesis, repeatedly performing a procedure yields experience (to the operating physician, surgical team, etc.), which in turn improves future outcomes. Under the selective-referral mechanism, better outcomes attract more patients. Practice makes perfect supports centralization, whereas selective referral does not.

If one could construct a computationally feasible instrument for hospital volume using exogenous variation, then the direction of causality could be investigated. One commonly used instrument for hospital volume is distance to hospital, e.g., the distance between the patient's and hospital's zip codes. For emergent conditions such as a heart attack, distance to hospital is likely to be a significant predictor of hospital volume, which is likely uncorrelated with patient outcomes [23, 35].

However, most admissions for cancer surgery are non-urgent. Patients need not go to the nearest hospital, but consider other factors like hospital quality in their hospital choice. Thus, distance to hospital is unlikely to be a good predictor of hospital volume. Two previous studies involving elective surgery use distance to hospital as an instrument for hospital volume on market concentration (derived from volume measures) and conclude that the instrumental variables analysis yields similar results to analyses without an instrument for volume [26, 36]. Further progress in testing the practice-makes-perfect hypothesis awaits identification of a valid instrument for volume in non-urgent procedures. Future studies should also consider using a measure of specialization (the number of a specific operation performed as a percentage of total hospital admissions) as a determinant of patient outcomes [37].

In addition to providing an analysis of six cancer procedures, this study highlights the benefits of (i) long-time-frame panel data and (ii) the appropriate use of fixed-effects and random-effects frameworks in controlling for unobserved heterogeneity. Application of the most appropriate statistical model improves the chances that referrals to high-volume centers will be made only when doing so improves patient outcomes.

Acknowledgments This study was a mutual collaboration of all authors. Vivian Ho acts as guarantor.

Compliance with Ethical Standards

Funding Support for this study was provided from a grant by Cancer Prevention and Research Institute of Texas (RP101207).

Conflict of interest Authors Woohyeon Kim, Stephen Wolff, and Vivian Ho declare that they have no conflicts of interest.

Statement of human rights For this type of study, formal consent is not required.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- 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(6):511–20. doi:10.7326/0003-4819-137-6-200209170-00012.
- The Leapfrog Group. Factsheet: Evidence-Based Hospital Referral. 2011. http://www.leapfroggroup.org/media/file/FactSheet_EBHR.pdf. Accessed 7 Mar 2016.
- Ross JS, Normand ST, Wang Y, Ko DT, Chen J, Drye EE, et al. Hospital volume and 30-day mortality for three common medical conditions. *N Engl J Med.* 2010;362:1110–8. doi:10.1056/NEJMs0907130.
- Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. *N Engl J Med.* 2011;364:2128–37. doi:10.1056/NEJMs1010705.
- Tsai TC, Joynt KE, Orav EJ, Gawande AA, Jha AK. Variation in surgical-readmission rates and quality of hospital care. *N Engl J Med.* 2013;369:1134–42. doi:10.1056/NEJMs1303118.
- American Cancer Society. Cancer Facts and Figures 2014. <http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2014>. Accessed 7 Mar 2016.
- Balzano G, Zerbi A, Capretti G, Rocchetti S, Capitanio V, Di Carlo V. Effect of hospital volume on outcome of pancreaticoduodenectomy in Italy. *Br J Surg.* 2008;95:357–62. doi:10.1002/bjs.5982.
- de Wilde RF, Besselink MGH, van der Tweel I, de Hingh IH, Van Eijck CH, Dejong CH, et al. Impact of nationwide centralization of pancreaticoduodenectomy on hospital mortality. *Br J Surg.* 2012;99:404–10. doi:10.1002/bjs.8664.
- Ra J, Paulson EC, Kucharczuk J, Armstrong K, Wirtalla C, Rapaport-Kelz R, et al. Postoperative mortality after esophagectomy for cancer: development of a preoperative risk prediction model. *Ann Surg Oncol.* 2008;15(6):1577–84. doi:10.1245/s10434-008-9867-4.
- Wouters MW, Karim-Kos HE, le Cessie S, Wijnhoven BP, Stassen LP, Steup WH, et al. Centralization of esophageal cancer surgery: does it improve clinical outcome? *Ann Surg Oncol.* 2009;16(7):1789–98. doi:10.1245/s10434-009-0458-9.
- LaPar DJ, Kron IL, Jones DR, Stukenborg GJ, Kozower BD. Hospital procedure volume should not be used as a measure of surgical quality. *Ann Surg.* 2012;256(4):606–15. doi:10.1097/SLA.0b013e31826b4be6.
- Miyata H, Motomura N, Ueda Y, Matsuda H, Takamoto S. Effect of procedural volume on outcome of coronary artery bypass graft surgery in Japan: implication toward public reporting and minimal volume standards. *J Thorac Cardiovasc Surg.* 2008;135(6):1306–12. doi:10.1016/j.jtcvs.2007.10.079.
- Miyata H, Motomura N, Ueda Y, Tsukihara H, Tabayashi K, Takamoto S. Toward quality improvement of thoracic aortic surgery: estimating volume-outcome effect from nationwide survey. *Eur J Cardiothorac Surg.* 2009;36:517–23. doi:10.1016/j.ejcts.2009.03.020.
- Bucher BT, Guth RM, Saito JM, Najaf T, Warner BW. Impact of hospital volume on in-hospital mortality of infants undergoing repair of congenital diaphragmatic hernia. *Ann Surg.* 2010;252(4):635–42. doi:10.1097/SLA.0b013e3181f5b538.
- Silber JH, Rosenbaum PR, Brachet TJ, Ross RN, Bressler LJ, Even-Shoshan O, et al. The hospital compare mortality model and the volume-outcome relationship. *Health Serv Res.* 2010;45:1148–67. doi:10.1111/j.1475-6773.2010.01130.x.
- Tepas JJ 3rd, Pracht EE, Orban BL, Flint LM. High-volume trauma centers have better outcomes treating traumatic brain injury. *J Trauma Acute Care Surg.* 2013;74(1):143–7. doi:10.1097/TA.0b013e3182788b5a.
- Allison PD. Fixed effects regression models. Los Angeles, CA: Sage Publications, Inc.; 2009.
- Sfekas A. Learning, forgetting, and hospital quality: an empirical analysis of cardiac procedures in Maryland and Arizona. *Health Econ.* 2009;18:697–711. doi:10.1002/hec.1400.
- Ho V, Short MN, Ku-Goto M. Can centralization of cancer surgery improve social welfare? *Forum Health Econ Policy.* 2012;15(3):107–31. doi:10.1515/fhep-2012-0016.
- Wei P, Liu S, Keller JJ, Lin H. Volume-outcome relation for acute appendicitis: evidence from a nationwide population-based study. *PLoS One.* 2012; doi:10.1371/journal.pone.0052539.
- Xie J, Dow WH. Longitudinal Study of child immunization determinants in China. *Soc Sci Med.* 2005;61:601–11. doi:10.1016/j.socscimed.2004.12.016.
- Hamilton BH, Hamilton VH. Estimating surgical volume-outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Econ.* 1997;6:383–95. doi:10.1002/(SICI)1099-1050(199707)6:4<383::AID-HEC278>3.0.CO;2-L.
- Kessler DP, McClellan MB. Is hospital competition socially wasteful? *Q J Econ.* 2000;115(2):577–615. doi:10.1162/00335300554863.
- Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. *N Engl J Med.* 1979;301(25):1364–9. doi:10.1056/NEJM197912203012503.
- Huesch MD. Learning by doing, scale effects, or neither? Cardiac surgeons after residency. *Health Serv Res.* 2009;44(6):1960–82. doi:10.1111/j.1475-6773.2009.01018.x.
- Gaynor M, Seider H, Vogt WB. The volume-outcome effect, scale economies, and learning-by-doing. *Am Econ Rev.* 2005;95(2):243–7. doi:10.1257/000282805774670329.

27. Ho V. Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. *J Health Econ.* 2002;21(5):873–85. doi:[10.1016/S0167-6296\(02\)00057-7](https://doi.org/10.1016/S0167-6296(02)00057-7).
28. Institute of Medicine (US) and National Research Council (US) National Cancer Policy Board; Hewitt M, Petitti D, editors. Interpreting the volume–outcome relationship in the context of cancer care. Washington (DC): National Academies Press (US); 2001.
29. Ho V, Aloia T. Hospital volume, surgeon volume, and patient costs for cancer surgery. *Med Care.* 2008;46:718–25. doi:[10.1097/MLR.0b013e3181653d6b](https://doi.org/10.1097/MLR.0b013e3181653d6b).
30. Chamberlain G. Analysis of covariance with qualitative data. *Rev Econ Stud.* 1980;47:225–38. doi:[10.2307/2297110](https://doi.org/10.2307/2297110).
31. Greene WH. *Econometric analysis*, 7th ed. New York: Pearson; 2012.
32. Schielzeth H, Nakagawa S. Nested by design: model fitting and interpretation in a mixed model era. *Methods Ecol Evol.* 2013;4(1):14–24. doi:[10.1111/j.2041-210x.2012.00251.x](https://doi.org/10.1111/j.2041-210x.2012.00251.x).
33. Cameron CA, Trivedi PK. *Microeconometrics: methods and applications*, 1st ed. England: Cambridge University Press; 2005.
34. Mundlak Y. On the pooling of time series and cross section data. *Econometrica.* 1978;46(1):69–85. doi:[10.2307/1913646](https://doi.org/10.2307/1913646).
35. Tsai AC, Votruba M, Bridges JFP, Cebul RD. Overcoming bias in estimating the volume-outcome relationship. *Health Serv Res.* 2006;41(1):252–64. doi:[10.1111/j.1475-6773.2005.00461.x](https://doi.org/10.1111/j.1475-6773.2005.00461.x).
36. Ho V, Town RJ, Heslin MJ. Regionalization versus competition in complex cancer surgery. *Health Econ Policy Law.* 2007;2:51–71.
37. Lee KCL, Sethuraman K, Yong J. On the hospital volume and outcome relationship: does specialization matter more than volume? *Health Serv Res.* 2015;50(6):2019–36. doi:[10.1111/1475-6773.12302](https://doi.org/10.1111/1475-6773.12302).