HEALTH ECONOMICS
Health Econ. **20**: 505–518 (2011)
Published online 11 April 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.1606

SIMULTANEOUS RELATIONSHIPS BETWEEN PROCEDURE VOLUME AND MORTALITY: DO THEY BIAS STUDIES OF MORTALITY AT SPECIALTY HOSPITALS?

DAVID BARKER[a,*], GARY ROSENTHAL[b,c] and PETER CRAM[b,c]

[a]*Tippie College of Business, University of Iowa, USA*
[b]*Center for Research in the Implementation of Innovative Strategies in Practice (CRIISP), Iowa City, VA Medical Center, USA*
[c]*Division of General Internal Medicine, Carver College of Medicine, University of Iowa, USA*

SUMMARY

Specialty hospitals have lower mortality rates for cardiac revascularization than general hospitals, but previous studies have found that this advantage disappears after adjusting for patient characteristics and hospital procedural volume. Questions have been raised about whether simultaneous relationships between volume and mortality might have biased these analyses. We use two-stage least squares with Hospital Quality Alliance scores and estimated market size as instruments for mortality and volume to control for possible simultaneity. After this adjustment, it is still the case that specialty hospitals do not have an advantage over general hospitals in mortality rates after cardiac revascularization. We find evidence of simultaneity in the relationship between volume and mortality. Copyright © 2010 John Wiley & Sons, Ltd.

Received 7 February 2008; Revised 18 December 2009; Accepted 16 February 2010

KEY WORDS: volume; mortality; hospitals; simultaneity

1. INTRODUCTION

The inverse relationship between hospital volume and patient mortality has been the subject of hundreds of published studies over the past 25 years. Although the nature of the relationship has been disputed (Vetrhus and Sondenaa, 2006; Sheikh, 2003; Sowden and Sheldon, 1998), a preponderance of studies concludes that high volume hospitals tend to have lower patient mortality (Chowdhury *et al.*, 2007; Birkmeyer *et al.*, 2002). Payors and policymakers have responded to these results by calling for the establishment of minimum volume thresholds for hospitals performing certain high-risk conditions and procedures (Shahian and Normand, 2003).

Most research in this area tests the assumption that high volume improves outcomes (the 'practice makes perfect' hypothesis). Another possibility, however, is that outcomes affect volume (the 'selective referral' hypothesis). In other words, physicians and patients might observe quality differences between hospitals and make choices that result in higher volume for high quality facilities. Both hypotheses might be simultaneously true, which would mean that volume and outcome are not exogenous variables, but are endogenously determined. Testing one of these hypotheses and ignoring the other can lead to biased and inconsistent results.

The possibility that the relationships between volume and outcome are simultaneously determined has important policy implications. If high hospital volume reduces patient mortality then a case can be

*Correspondence to: Tippie College of Business, University of Iowa, 519 S. Summit Street, Iowa City, IA 52240, USA. E-mail: david-barker@uiowa.edu

made for closing small hospitals and regionalizing care (Shahian and Normand, 2003; Birkmeyer *et al.*, 2001, 2002). On the contrary, if 'selective referral' explains the relationship between volume and outcome, then regionalization could actually worsen outcomes if the wrong hospitals are chosen for expansion. The benefits of regionalization must also be weighed against the cost of transporting patients over long distances (Birkmeyer *et al.*, 2003), and hence, knowing the magnitude of the 'practice makes perfect' effect is important. Failure to take account of the 'selective referral' effect will tend to exaggerate the estimated 'practice makes perfect' effect, biasing policy recommendations toward regionalization.

Another reason that understanding the nature of the relationship between volume and outcome is important is that hospital volume is often used as a control variable in studies of a variety of topics. As the correlation of volume and outcome is very well documented, studies that examine the effect of any factor on outcomes must control for hospital volume. If the simultaneous relationships between volume and outcome are not taken into account, then the results of these studies may be unreliable.[1]

We use the case of specialty cardiac hospitals in comparison with general hospitals to explore the potential effects of simultaneity in the estimation of the volume–outcome relationship and the use of volume as a control variable. In a recent study, Cram *et al.* (2005) found that patient mortality after cardiac revascularization is lower in specialty cardiac hospitals than general hospitals even after adjusting for patient characteristics. However, after adjusting for hospital procedural volume, this advantage becomes statistically insignificant, leading the authors to conclude that there is no inherent benefit to hospital specialization above and beyond the benefits offered by greater volume. Ballard (2005) responded to these results by commenting that adjustment for volume is 'inappropriate… given its role in the causal pathway related to hospital quality and efficiency.'

Ballard is correct in pointing out that simultaneity between procedural volume and mortality could have biased the results of Cram *et al.* (2005). A regression, logistic or otherwise, with mortality as the dependent variable and volume as an independent variable that fails to take account of both relationships could be biased toward finding either an advantage or a disadvantage for specialty hospitals over general hospitals. The direction of the bias will depend on the relative volatility of factors influencing hospital quality and procedure volume.

In Section 2, we briefly review the literature on the relationship between volume and outcome. In Section 3, we explain why regression analysis of the relationship between mortality and procedure volume that does not take account of simultaneity may be biased, and the ability of two-stage least squares to correct this bias. In Section 4, we apply two-stage least squares to the Medicare administrative data used by Cram *et al.* (2005) to compare specialty and general hospitals to examine how application of this method might change the study findings. Specification and robustness tests are presented in Section 5, and Section 6 concludes the study.

2. LITERATURE REVIEW

The pioneering study of the relationship between volume and outcome (Luft *et al.*, 1979), inspired by the economics literature on industrial learning curves, reported a strong correlation between volume and outcome. It mentioned, however, the possibility that the causal relationship might be from outcome to volume, rather than the other way around. In a second study, Luft (1980) explored the directionality of this relationship using the econometric technique of two-stage least squares and confirmed this concern. He found evidence suggesting that the effect of outcome on volume (the 'selective referral' hypothesis) was stronger than the effect of volume on outcome (the 'practice makes perfect' hypothesis).

A later study (Hughes *et al.*, 1988) found evidence that the relationship was strong in both directions. Another study (Luft *et al.*, 1990) found that hospital quality affected patient's hospital choice even

[1]There are hundreds of studies to which this criticism applies. See, for example, Birkmeyer *et al.* (2002) and Begg *et al.* (2002).

DOI: 10.1002/hec

before explicit data on quality were available, supporting the relevance of the 'selective referral' hypothesis.

Although failure to take account of both hypotheses has the potential to bias the results of studies of hospital volume, very few studies in mainstream medical research journals demonstrate awareness of the bi-directional volume–outcome relationship. Some have addressed the issue using longitudinal data. Hannan *et al*. (1992, 1995) found evidence supporting both directions of causality. Hamilton and Ho (1998) and Hamilton and Hamilton (1997) found no evidence supporting the 'practice makes perfect' hypothesis after controlling for fixed quality effects, supporting the 'selective referral' hypothesis.

Studies using cross-sectional data and correcting for possible simultaneity are even less common than longitudinal studies. Other than the work of Harold Luft, we are aware of only three studies that explicitly test or correct for simultaneity. Farley and Ozminkowski (1992), using longitudinal data and two-stage least squares, found that the 'practice makes perfect' effect disappeared for coronary artery bypass grafting (CABG) once the 'selective referral' effect was controlled for, although the effect remained for some other procedures. Norton *et al*. (1998) found a *t*-statistic of only 1.80 in a second-stage regression and concluded that they were unable to reject exogeneity, and hence a correction for simultaneity was not required. This conclusion was criticized as premature by Luft (1998a,b), who argued that a *t*-value that approached significance should have led the authors to explore additional model specifications to rule out endogeneity. Tsai *et al*. (2006) used patient distance from hospitals to construct an instrument for volume. After adjusting for simultaneity, the 'practice makes perfect' effect became statistically insignificant.

Although 'selective referral' is ignored in the vast majority of studies of volume and outcome, there is considerable evidence that it should be taken into account. Failure to take simultaneity into account does not, of course, call into question the correlation between volume and outcome. As Luft (2003) states, if he were injured in an unfamiliar city, he 'would ask the ambulance driver to avoid taking me to the low-volume facility.' However, correlation does not imply causation; hence policymakers deciding whether to regionalize care require more sophisticated statistical analysis than do patients deciding where to obtain care.

Although there are few studies that take account of simultaneity in the analysis of the relationship between volume and outcome itself, there is even less appreciation of the statistical difficulties that simultaneity imposes on the use of hospital volume as a control variable in a wide variety of situations.

3. SIMULTANEITY BIAS

Suppose that the mortality rate at a hospital, $\mathrm{DR_h}$, is a linear function of three variables: procedure volume, V_h, whether the hospital is a specialty or a general hospital, S_h, and hospital quality, $\mathrm{HQ_h}$. In addition, procedure volume is a linear function of $\mathrm{DR_h}$, S_h, and the local market size, $\mathrm{MS_h}$. $\mathrm{HQ_h}$ is some indicator of the quality of care given at a hospital that influences mortality rates, and $\mathrm{MS_h}$ is a measure of the ability of a hospital to achieve a high volume of procedures, such as the population of the city in which the hospital is located.

$$
\begin{aligned}
\text{Mortality:}\quad & \mathrm{DR_h} = \alpha_1 + \alpha_2 V_\mathrm{h} + \alpha_3 \mathrm{HQ_h} + \alpha_4 S_\mathrm{h} + \varepsilon_{1\mathrm{h}} \\
\text{Volume:}\quad & V_\mathrm{h} = \beta_1 + \beta_2 \mathrm{DR_h} + \beta_3 \mathrm{MS_h} + \beta_4 S_\mathrm{h} + \varepsilon_{2\mathrm{h}}
\end{aligned}
\tag{1}
$$

The first equation in Equation (1) represents the improvement in mortality that results from learning by doing (the 'practice makes perfect' hypothesis); when more procedures are performed at a hospital the quality of the procedures might improve and mortality would fall. Higher values of the independent measure of hospital quality, $\mathrm{HQ_h}$, also reduce mortality. The second equation represents the effect of better quality of care on volume (the 'selective referral' hypothesis); if hospitals are able to achieve low mortality, they attract more patients. Greater market size, $\mathrm{MS_h}$, also increases volume, independent of

DOI: 10.1002/hec

the quality of care. The model allows the possibility that being a specialty hospital affects the quality of care and the volume of procedures, independent of the other variables.

Ordinary least squares and logistic regression require that the independent variables be uncorrelated with the residual term in order for parameter estimates to be unbiased. The mortality equation above shows that DR_h is correlated with the error term ε_{1h}. The volume equation shows that V_h is correlated with DR_h. This means, however, that V_h is correlated with ε_{1h}. As V_{1h} is an independent variable in the mortality equation, the requirement that independent variables be uncorrelated with the residual term is violated. As a result, all of the parameter estimates will be biased, including α_4 and β_4, the parameters of interest.

If the 'practice makes perfect' equation is estimated alone, the direction of bias will be unknown, as information about the 'selective referral' effect will not be available. Depending on the relative variation of hospital quality and hospital market size, the regression might show specialty hospitals to have lower or higher mortality than general hospitals, even if there is actually no difference between them. It is also possible that the regression will show no difference between specialty and general hospitals, even if one has a clear advantage in mortality.

Two-stage least squares estimation using market size and hospital quality as instrumental variables should correctly estimate the parameters of both relationships.

4. EMPIRICAL RESULTS

To test whether simultaneity bias is important, we replicated analyses conducted by Cram *et al.* (2005), and then checked to see whether two-stage least squares analysis would change the results. We used a sample of 335 255 Medicare patients who underwent CABG in 2000 and 2001. A logistic regression was used to estimate the probability of death within 30 days of undergoing CABG based on the same set of patient characteristics that investigators have used previously (Cram *et al.*, 2005). For each hospital, the risk-adjusted mortality rate was calculated as the difference between the hospital's actual mortality rate and the mean of the probability of death estimated by the logistic regression for the hospital's patients. Hospitals were classified as specialty cardiac or general hospitals using methods that have been described previously (Cram *et al.*, 2005). Procedure volume for each hospital was the total number of CABG procedures performed in 2000 and 2001.

To correct for simultaneity bias using two-stage least squares, instruments for hospital quality and volume are needed. A measure of hospital quality is available from the Hospital Quality Alliance (HQA) program,[2] which reports several measures of whether hospitals follow accepted practices in their management of certain medical conditions. HQA reports the percentage of the time each hospital follows these practices. A recent study by Jha *et al.* (2007) demonstrates that these scores are inversely correlated with mortality. We calculated an average score for each hospital using the average of the available scores for each hospital on heart-related issues.[3] Scores of zero on any quality measure were eliminated. Only accredited hospitals were included, as they are the only hospitals that are required to report HQA data.

The total number of hospitals in the Medicare data was 1040. Of these hospitals, 48 were not included in the HQA database and were eliminated, leaving 992. One hospital was in the HQA database but did not report any data. The number of specialty hospitals in this sample was 27.

The ratio of nurses employed by the hospital to total admissions and whether the hospital was for-profit or not were also used as instruments for hospital quality. These data were obtained from a

[2]Available from http://www.cms.hhs.gov/HospitalQualityInits/25_HospitalCompare.asp.
[3]Results were similar using the average score on all the issues.

DOI: 10.1002/hec

Table I. Variable definitions

Variable	Definition
DR	Difference of hospital mortality rate and prediction from patient health
V	Number of procedures performed at the hospital in 2000 and 2001
S	Equal to one if hospital is a specialty hospital, zero otherwise
HQ	Mean score on HQA survey on heart related issues
NURSE	Ratio of nurses employed to hospital admissions
PROF	One if hospital is for-profit, zero otherwise
RACE	Percentage of hospital's patients that are African–American
POP	Population in hospital MSA over 65 years of age
NHOS	Number of hospitals in MSA
EVOL	Expected volume in hospital based on geographic distribution of patients

survey conducted by the American Hospital Association. Fifty-eight hospitals in our sample did not report nursing data.

As quality of care may be affected by the racial identity of patients, the percentage of patients who were African–American was also used as an instrument in some regressions. Disparities by race in the quality of care are reported in the National Healthcare Disparities Report (Agency for Healthcare Research and Quality, 2007). Racial differences in the acceptance of patients by cardiac specialty hospitals are discussed in Nallamothu *et al*. (2008).

We constructed an instrument for hospital volume using hospital locations and patient zipcodes. The distance from the centroid of each patient zipcode to each hospital was calculated.[4] We assume that the probability that a patient goes to a particular hospital varies with the reciprocal of the squared distance. A hospital's expected volume is, therefore, the sum over patients of the reciprocal of the squared distance divided by the sum of this measure across hospitals, as shown in Equation (2). EV_j is the instrument, or the expected volume, of hospital j, P is the total number of patients, H is the total number of hospitals, and d_{ij} is the distance between patient i and hospital j.

$$\mathrm{EV}_j = \sum_{i=1}^{P} \frac{1/d_{ij}^2}{\sum_{j=1}^{H}(1/d_{ij}^2)} \tag{2}$$

Other instruments for hospital volume were the population over age 65 of the MSA in which the hospital is located, and the number of hospitals in the MSA.

Variable definitions are summarized in Table I, and summary statistics are summarized in Table II.

Ordinary least squares results for mortality are summarized in Table III. All four models shown include S, NURSE, PROF, and HQ. Model 1 includes volume, while model 2 does not. Models 3 and 4 also include the instruments for volume, EVOL, NHOS, and POP. Model 3 includes volume, while model 4 does not.

S, the specialty hospital indicator, is not statistically significant in any of the models. Volume is statistically significant in model 1, but its inclusion does not have much effect on the statistical significance of the other variables, except that NURSE becomes statistically significant when volume is removed. The inclusion of volume slightly reduces the statistical significance of HQ, but it remains statistically significant at the 5% level.

EVOL, the expected volume calculated from the distance of patients from the hospital, is an even stronger predictor of mortality than volume itself. It is possible that our measure of volume in 2000 and 2001 represents a transitory measure of volume, while EVOL better reflects long-term volume. For-profit status predicts higher mortality in all model specifications.

[4]The patients and hospitals are spread across the United States, and hence, in some cases, the curvature of the earth is a significant factor. We use a formula for distance that takes this into account.

DOI: 10.1002/hec

Table II. Summary statistics

Variable	N	Mean	Std. Dev	Min	Max
General hospitals					
DR	965	0.004	0.028	−0.075	0.177
V	965	322.2	304.968	2	2174
HQ	964	0.853	0.06	0.5	0.979
NURSE	918	0.088	0.053	0	0.672
PROF	965	0.153	0.361	0	1
RACE	965	0.054	0.101	0	0.933
POP	965	132 088	204 910	2276	998 963
NHOS	965	33.632	43.171	1	158
EVOL	965	315.558	131.783	9.742	958.931
Specialty hospitals					
DR	27	−0.002	0.015	−0.024	0.054
V	27	639.481	428.685	58	2252
HQ	27	0.886	0.061	0.701	0.983
NURSE	18	0.095	0.072	0.032	0.321
PROF	27	0.481	0.509	0	1
RACE	27	0.038	0.047	0	0.21
POP	27	132 667	117 536	2601	403 053
NHOS	27	49.407	56.159	4	158
EVOL	27	334.396	169.485	124.798	746.024

Table III. Ordinary least squares, mortality, and medicare data

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	0.03022	0.03355	0.02644	0.02886
	(0.03131)	(0.01704)	(0.06709)	(0.04466)
S	−0.00329	−0.00726	−0.0027	−0.0047
	(0.62655)	(0.27635)	(0.69164)	(0.48319)
NURSE	0.03272	0.03963	0.0285	0.02944
	(0.06436)	(0.02461)	(0.11632)	(0.10485)
PROF	0.00578	0.00691	0.00544	0.00586
	(0.02782)	(0.00823)	(0.03776)	(0.02465)
HQ	−0.03138	−0.03977	−0.02231	−0.02525
	(0.04986)	(0.01208)	(0.16945)	(0.11784)
V	−0.00001		−0.00001	
	(0.00205)		(0.11679)	
EVOL			−0.00002	−0.00002
			(0.03137)	(0.00108)
NHOS			−0.00002	−0.00002
			(0.39159)	(0.51375)
POP			0.00000	0.00000
			(0.05399)	(0.07106)
R-squared	0.03508	0.02515	0.04682	0.04428
Observations	935	935	935	935

Note: Each observation is a single hospital. The dependent variable is the difference between the observed mortality rate and the mortality rate predicted by a logistic regression of patient death on patient characteristics (DR). See Table I for variable definitions. P-values are in parentheses.

Table IV summarizes ordinary least squares results for volume. Here, the specialty hospital indicator is highly significant. Expected volume based on patient distance is also a strong predictor of volume. Holding population constant, a larger number of hospitals in an MSA predicts lower volume. Without including for profit status and hospital quality, higher mortality rates predict lower volume, although this relationship is not statistically significant if hospital quality is included.

Two-stage least squares results are reported in Table V. In model 1, hospital quality is used as an instrument for mortality, but not for volume. In model 2, hospital quality appears in both equations.

DOI: 10.1002/hec

Table IV. Ordinary least squares, volume, and medicare data

Variable	Model 1	Model 2
Intercept	−30.8428	−428.5591
	(0.2029)	(0.0016)
S	303.3737	363.339
	(<0.0001)	(<0.0001)
DR	−633.1082	−486.5203
	(0.0399)	(0.1168)
EVOL	1.1817	1.1026
	(<0.0001)	(<0.0001)
NHOSP	−0.7646	−1.1269
	(0.0032)	(<0.0001)
POP	0.0001	0.0001
	(0.2683)	(0.0108)
NURSE		−158.0726
		(0.3565)
PROF		−73.9483
		(0.0028)
HQ		526.2584
		(0.0006)
R Squared	0.293	0.317
Observations	992	935

Note: Each observation is a single hospital. The dependent variable is the number of procedures performed at the hospital in 2000 and 2001 (V). See Table I for variable definitions. P-values are in parentheses.

In model 3, hospital quality appears in both equations, but patient race is used as an instrument for quality instead of nursing care. Eliminating the nursing variable adds 56 hospitals that did not have data on the size of the nursing staff. In model 4, hospital quality is also eliminated, and race is included.

The 'selective referral' effect is statistically significant in all four model specifications, although including hospital quality in the volume equation reduces the statistical significance of the mortality rate to barely the 5% level. In the other three model specifications, the 'selective referral' effect is highly statistically significant ($P<0.0039$). The average parameter estimate across the models implies that an improvement of 1% in a hospital's adjusted mortality rate would attract 127 additional CABG patients. The 'practice makes perfect' effect is also statistically significant in all four-model specifications.[5]

Specialty hospital status does not appear to have an independent effect on mortality ($P<0.2439$). While not statistically significant, the estimated difference in mortality between specialty and general hospitals without any controls is non-trivial. The difference in mean mortality rates summarized in Table II is 0.006 (0.6%). Extrapolating this difference to the estimated 300 000 CABG surgeries performed on Medicare beneficiaries in our study hospitals during 2000 and 2001 would translate into 2000 lives being saved by switching all patients to specialty hospitals. Even after controlling for various hospital characteristics the mortality advantage in specialty hospitals is maintained. Although statistically insignificant, in our final two-stage least squares results, the estimated size of specialty status remains close to 0.006.

Our finding of a large effect that is statistically insignificant produces a challenge for policy makers. The magnitude of the coefficient on specialty hospital status suggests a large effect, but the lack of statistical significance suggests that there is a high probability that the effect would not be replicated in another sample.

[5]The parameter estimates are consistent with equilibrium in a dynamic version of the model. Equilibrium is consistent with the reciprocal of the coefficient on DR in the volume equation being larger than the coefficient on V in the mortality equation.

DOI: 10.1002/hec

Table V. Two-stage least squares and medicare data

Variable	Model 1	Model 2	Model 3	Model4
Mortality (DR)				
Intercept	0.0236	0.0236	0.0206	0.0112
	(0.1013)	(0.1013)	(0.1314)	(0)
V	−0.000023	−0.000023	−0.000026	−0.000028
	(0.0003)	(0.0003)	(<0.0001)	(<0.0001)
S	−0.0058	−0.0058	−0.0055	−0.0064
	(0.3864)	(0.3864)	(0.3222)	(0.2439)
PROF	0.0045	0.0045	0.0026	0.0029
	(0.0982)	(0.0982)	(0.2946)	(0.2463)
RACE			0.0236	0.0245
			(0.0097)	(0.0069)
HQ	−0.0173	−0.0173	−0.0113	
	(0.3133)	(0.3133)	(0.4921)	
NURSE	0.0241	0.0241		
	(0.1874)	(0.1874)		
R-squared	0.03802	0.03802	0.04820	0.04631
Observations	935	935	991	992
Volume (V)				
Variable	Estimate			
Intercept	146.56	−95.3232	−35.6173	118.7428
	(0.0482)	(0.7594)	(0.889)	(0.0389)
DR	−14 314.596	−11 170.184	−12 315.628	−13 039.259
	(0.0039)	(0.0536)	(0.0012)	(0.0002)
S	371.9719	362.5582	299.3409	314.8988
	(0.0008)	(0.0002)	(0.0003)	(0.0002)
POP	0.0003	0.0003	0.0002	0.0002
	(0.0147)	(0.0133)	(0.0292)	(0.0376)
NHOSP	−1.3413	−1.2935	−0.9227	−0.9325
	(0.0051)	(0.0017)	(0.0241)	(0.0278)
EVOL	0.7623	0.8386	0.8207	0.8101
	(0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
HQ		242.674	173.4251	
		(0.4275)	(0.5186)	
R-squared	0.12667	0.16727	0.15589	0.14439
Observations	935	935	991	992

Note: Each observation is a single hospital. The dependent variable is the ratio of the observed mortality rate to the mortality rate predicted by a logistic regression of patient death on patient characteristics. The first panel shows the results of estimating mortality as a function of volume using the 'practice makes perfect' equation, and the second panel shows the results of estimating volume as a function of mortality using the 'selective referral' equation. See Table I for variable definitions. *P*-values are in parentheses.

Specialty hospitals have higher volume than general hospitals ($P \leq 0.0008$). The ratio of the size of the nursing staff to total admissions ($P = 0.1874$) and whether the hospital is for profit ($P > 0.0982$) are not statistically significant in the mortality equation. The population of the MSA in which the hospital is located has a statically significant positive effect on procedure volume ($P < 0.0376$), as does the expected volume based on patient and hospital locations ($P \leq 0.0001$). The number of hospitals in the MSA is also statistically significant ($P \leq 0.0278$).

Hospital quality, as measured by HQA scores, does not have a statistically significant effect on mortality ($P \geq 0.3133$) in these regressions. This may, however, be due to the fact that some hospital quality scores are based on a small number of observations. We ranked the hospitals by the total number of observations in their quality surveys and eliminated the bottom third. Table VI reports the results of two-stage least squares estimation using this sample. The results are essentially unchanged, except that HQ becomes statistically significant in the mortality regression ($P < 0.0372$). If race is not included, HQ is more statistically significant ($P = 0.0169$).

DOI: 10.1002/hec

Table VI. Two-stage least squares and medicare data

Variable	Model 1	Model 2	Model 3	Model 4
Mortality (DR)				
Intercept	0.051	0.051	0.0458	0.0125
	(0.0018)	(0.0018)	(0.0039)	(0)
V	−0.000026	−0.000026	−0.000027	−0.000030
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
S	−0.0066	−0.0066	−0.0073	−0.0089
	(0.3544)	(0.3544)	(0.2812)	(0.1916)
PROF	0.0046	0.0046	0.0037	0.0047
	(0.129)	(0.129)	(0.196)	(0.1027)
RACE			0.0185	0.0222
			(0.111)	(0.057)
HQ	−0.046	−0.046	−0.0394	
	(0.0169)	(0.0169)	(0.0372)	
NURSE	0.0087	0.0087		
	(0.6677)	(0.6677)		
R-squared	0.07228	0.07228	0.07916	0.06585
Observations	629	629	663	663
Volume (V)				
Variable	Estimate			
Intercept	154.0106	577.7891	520.1725	190.3376
	(0.0253)	(0.4414)	(0.3096)	(0.0177)
DR	−10 546.26	−15 268.774	−15 041.335	−13 851.497
	(0.0031)	(0.1027)	(0.0102)	(0.0016)
S	406.1595	417.1442	385.0556	399.7175
	(0.0003)	(0.0023)	(0.0025)	(0.001)
POP	0.0002	0.0002	0.0001	0.0001
	(0.0722)	(0.112)	(0.287)	(0.2403)
NHOSP	−1.3673	−1.4545	−0.974	−0.9653
	(0.0052)	(0.0168)	(0.0781)	(0.0659)
EVOL	0.8211	0.6719	0.6921	0.7122
	(0)	(0.0459)	(0.0028)	(0.0005)
HQ		−419.9916	−370.2452	
		(0.57)	(0.4783)	
R-squared	0.19298	0.14214	0.14834	0.15805
Observations	629	629	663	663

Note: Each observation is a single hospital. Only hospitals with large sample sizes for measuring quality are included. The dependent variable is the ratio of the observed mortality rate to the mortality rate predicted by a logistic regression of patient death on patient characteristics. The first panel shows the results of estimating mortality as a function of volume using the 'practice makes perfect' equation, and the second panel shows the results of estimating volume as a function of mortality using the 'selective referral' equation. See Table I for variable definitions. *P*-values are in parentheses.

Although specialty hospital status does not appear to have a statistically significant direct effect on mortality, it does affect mortality indirectly through volume. Specialty hospitals have higher volume than general hospitals, and higher volume is associated with lower mortality. Solving equation system (1) for mortality and taking the derivative of mortality with respect to specialty status yields the following:

$$\frac{\mathrm{dDR_h}}{\mathrm{d}V_\mathrm{h}} = \frac{\alpha_2\beta_4+\alpha_4}{1-\alpha_2\beta_2} \tag{3}$$

Our estimates of the parameters α_2, α_4, β_2, and β_4 are themselves normally distributed random variables, but the expression in Equation (3), a function of these normally distributed variables, is not normally distributed. To estimate the probability distribution of this expression, we use a bootstrap method, which simulates the existence of many random samples of hospitals. For each simulated sample, we randomly selected observations with replacement, meaning that a particular hospital might

DOI: 10.1002/hec

Table VII. Bootstrap estimate of total effect of specialty status on mortality

	Model 1	Model 2	Model 3	Model 4
All hospitals				
Median	−0.0191	−0.0173	−0.0190	−0.2270
95th percentile	−0.0087	−0.0076	−0.0069	−0.0094
5th percentile	−0.0405	−0.0356	−0.0425	−0.0526
Fraction positive	0.0077	0.0094	0.0160	0.0200
Hospitals used in Table VI				
Median	−0.022	−0.0222	−0.0271	−0.0329
95th percentile	−0.0114	−0.0104	−0.0117	−0.0177
5th percentile	−0.0422	−0.0566	0.0757	−0.0838
Fraction positive	0.0009	0.0200	0.0341	0.0188

Descriptive statistics of values obtained bootstrapping the sample and calculating the expression in Equation (3).

appear multiple times, or not at all. We constructed 10 000 samples in this way and repeated the two-stage least squares estimation for each sample. For each of the 10 000 resulting parameter estimates, we computed the expression in Equation (3). The resulting set of 10 000 values of the expression in Equation (3) is an approximation of the probability distribution of this expression.

As the mean and variance of the resulting non-normal distribution is not well defined, we examined the median and percentiles of the distribution, which are summarized in Table VII. The lowest value for the median is −1.9%, which is a very large effect. The effect is positive at least 97.6% of the time, and the 95th percentile estimates are all above 0.69%.

Specialty hospitals clearly have mortality rates that are significantly lower than general hospitals that are similar in all respects except volume. It is important to point out, however, that volume is the only factor contributing to this reduction in mortality. Controlling for volume, specialty hospitals have no statistically significant mortality advantage over general hospitals.

5. SPECIFICATION AND ROBUSTNESS TESTS

A Hausman–Wu test for endogeneity on this system produces a test statistic with a P-value below 0.001, which indicates that two-stage least squares estimation is preferable to OLS.[6] Augmented regression using the residuals from the first stage in the second-stage regression also indicates endogeneity with a high level of confidence.

The Sargan 'validity of instruments' test (Stewart and Gill, 1991) was used to check the appropriateness of the instrumental variables used in this analysis. The test statistic was well below the critical value for both equations, suggesting that the instruments are appropriate. The Sargan test involves regressing the residuals from the two-stage least squares estimates on the instrumental variables. The test statistic is calculated from the sum of the squared values of the residuals of these regressions. A high value of the test statistic indicates that the variables are not good choices for instruments because they are correlated with the error terms ε_1 and ε_2.

A potential problem with these results is that the HQA measure of hospital quality may be correlated with specialty status. If specialty hospitals offer higher quality care than general hospitals and the HQA measure captures this advantage, then including the HQA measure in a regression with a dummy variable indicating specialty status may reduce the observed statistical significance of specialty status. To investigate this possibility, we examined the correlation between the HQA score and specialty status. Specialty hospitals had an average HQA score of 0.8487, whereas general hospitals had an average score

[6]A similar test indicates that the use of three-stage least squares is unnecessary. Three-stage least squares is similar to two-stage least squares, but corrects for possible correlation between the errors of the two equations, ε_1 and ε_2.

DOI: 10.1002/hec

of 0.8401. This difference is not statistically significant ($P = 0.4893$). The correlation between the HQA score and the specialty status dummy variable is 0.025 and is not statistically different from zero. This finding is consistent with the results of Popescu *et al.* (2008), who found that quality measures were similar between cardiac specialty hospitals and general hospitals.

Another potential issue is the linearity of the relationship between volume and outcome. We included a term for the square of volume in the 'practice makes perfect' equation and found that the term was not statistically significant, and that including this term did not affect the magnitudes or statistical significance of the other variables. This result is consistent with the findings of Marcin *et al.* (2008) with regard to cardiac revascularization and those of Vakili *et al.* (2001) with regard to angioplasty. Neither study found evidence of non-linearity in the relationship between volume and outcome.

We also performed the estimation using the logarithm of volume as in Ho (2000). The basic results were unchanged, but the fit deteriorated. The R-squared measure of fit was lower using the logarithm of volume in every model specification and in both samples except for the mortality equation in model 1 using the larger sample, where R-squared was 0.03829 instead of 0.03802. (The mortality equation is the same in models 1 and 2.) Volume remained statistically significant in the mortality equation in each specification. Without transforming volume, mortality is statistically significant in each specification of the volume equation except model 2, but using the logarithm of volume, mortality is statistically significant using the larger sample ($P = 0.0241$) and using the smaller sample, it is closer to being statistically significant ($P = 0.0881$). If both excess mortality and volume are transformed by taking logarithms (adding a constant to excess mortality to avoid taking logarithms of negative numbers), R-squared is lower in every specification, equation, and sample except for the volume equation of model 2 using the smaller sample, where the R-squared increased from 0.14214 to 0.20055. This transformation did not change the statistical significance of volume in the mortality equations or mortality in the volume equations.

It is also possible that specialty status for a hospital and mortality might be endogenous. Specialization might be more profitable in locations where populations are wealthier and healthier; hence, conditions that cause low mortality might cause specialization, while specialization might also affect mortality. If this were the case, however, then we would expect specialty status to be statistically significant in the OLS regressions of mortality, and it is not. Although Table I indicates that mortality rates are lower for specialty than for general hospitals, the difference, even without any controls, is not statistically significant. If endogeneity of this sort were properly controlled for, it seems very likely that the statistical significance of specialty status would be reduced even further.

Results of regression analysis of HQA scores on hospital characteristics are summarized in Table VIII. Specialty status is not statistically significant in any of the regressions. Higher volume hospitals have higher HQA scores, whereas hospitals with greater market potential have lower scores. This result may indicate that hospitals in more competitive environments have higher HQA scores. Patient income is not statistically significant, but a higher percentage of patients who are black is associated with lower HQA scores.

6. CONCLUSION

Understanding the relationship between procedure volume and mortality is of vital importance, both for academic research and for health policy. As the relationship has been very well established, volume is commonly used as a control variable in many studies. Policymakers depend on accurate estimates of the relationship when they decide whether to concentrate medical facilities. Rarely, however, have any researchers taken account of the simultaneity of the relationship, which might seriously bias their results. Apart from the pioneering but largely overlooked work of Harold Luft on this subject 30 years

DOI: 10.1002/hec

Table VIII. Regression analysis of HQA scores

Variable	A	B	C
Intercept	0.85339	0.84378	0.85171
	(<0.0001)	(<0.0001)	(<0.0001)
S	0.03218	0.01834	0.01778
	(0.0059)	(0.1122)	(0.1175)
V		0.00004	0.00004
		(<0.0001)	(<0.0001)
MS		−0.00	−0.00
		(0.0417)	(0.0088)
INC			0.00
			(0.9386)
BLACK			−0.11052
			(<0.0001)
Observations	991	991	991
R-squared	0.0076	0.0591	0.0925

Each observation is a single hospital. The dependent variable is the average score a hospital received from the Hospital Quality Alliance. *S* is a dummy variable indicating whether a hospital is a specialty or a general hospital with a value of one indicating a specialty hospital. *V* is the number of CABG procedures performed at a hospital in 2000 and 2001. MS is the population over age 65 of the MSA the hospital is located in, divided by the number of hospitals in the MSA. INC is the average income of the zipcodes of patients at each hospital, weighted by the number of patients in each zipcode. BLACK is the fraction of the patients at a hospital undergoing CABG who are black. *P*-values are in parentheses.

ago, out of hundreds of studies on volume and outcome, ours is one of very few to examine the issue of simultaneity.

Procedure volume and mortality rates may be simultaneously related in two different ways. First, proficiency might improve as more procedures are carried out at a hospital, lowering mortality rates. Second, low mortality rates might attract patients, raising procedural volume for a hospital. If both of these relationships occur, then traditional methodologies for measuring either effect will produce biased and inconsistent results. Studies of topics related to these effects, such as mortality differences between specialty and general hospitals, may also have biased results.

In this study, we have demonstrated the possibility that previous work on specialty hospitals might have been biased due to failure to control for simultaneity in the relationship between procedure volume and mortality. After correcting for simultaneity, however, we confirm the result reported in Cram *et al.* (2005) that specialty cardiac hospitals have no mortality rate advantage over general hospitals with the same procedure volume.

More importantly, perhaps, we find evidence that mortality rates influence the number of patients that a hospital is able to attract, and therefore, call into question a significant body of literature proposing that hospitals can improve patient outcomes by simply increasing their volume. Future research should reevaluate these study, taking account of possible simultaneity bias.

ACKNOWLEDGEMENTS

Funding sources had no role in the analyses or drafting of this manuscript. None of the authors have any conflicts of interest. Dr. Rosenthal is the Director of the Center for Research in the Implementation of Innovative Strategies in Practice (CRIISP) at the Iowa City VA Medical Center, which is funded through the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service. Dr. Cram is supported by a K23 career development award (RR01997201) from the NCRR at the NIH and the Robert Wood Johnson Physician Faculty Scholars Program. This work is also funded by R01 HL085347-01A1 from NHLBI at the NIH. The views

DOI: 10.1002/hec

expressed in this study are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. We thank Kurt Anstreicher, David Ballard, John Geweke, Gene Savin, and Mary Vaughan-Sarrazin for their comments and assistance.

REFERENCES

Agency for Healthcare Research and Quality. 2007. *National Healthcare Disparities Report*. Rockville: Agency for Healthcare Research and Quality. Available from http://www.ahrq.gov/qual/nhdr07/nhdr07.htm.

Ballard D. 2005. Cardiac revascularization in specialty and general hospitals, Letter to the Editor. *The New England Journal of Medicine* **352**: 2754–2756.

Begg CB, Riedel ER, Bach PB *et al*. 2002. Variations in Morbidity after Radical Prostatectomy. *New England Journal of Medicine* **346**: 1138–1144.

Birkmeyer JD, Finlayson EV, Birkmeyer CM. 2001. Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. *Surgery* **130**: 415–422.

Birkmeyer JD, Siewers AE, Finlayson EVA, Stukel TA, Lucas FL, Batista I, Welch HG, Wennberg DE. 2002. Hospital volume and surgical mortality in the United States. *The New England Journal of Medicine* **346**: 1128–1137.

Birkmeyer JD, Siewers AE, Marth NJ, Goodman DC. 2003. Regionalization of high-risk surgery and implications for patient travel times. *Journal of the American Medical Association* **290**: 2703–2708.

Chowdhury MM, Dagash H, Pierro A. 2007. A systematic review of the impact of volume of surgery and specialization on patient outcome. *British Journal of Surgery* **94**: 145–161.

Cram P, Rosenthal G, Vaughan-Sarrazin M. 2005. Cardiac revascularization in specialty and general hospitals. *The New England Journal of Medicine* **352**: 1454–1462.

Farley DE, Ozminkowski RJ. 1992. Volume–outcome relationships and in hospital mortality: the effect of changes in volume over time. *Medical Care* **30**: 77–94.

Hamilton BH, Hamilton VH. 1997. Estimating surgical volume–outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Economics* **6**: 383–395.

Hamilton BH, Ho V. 1998. Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. *Medical Care* **36**: 892–903.

Hannan EL, Kilburn Jr H, O'Donnell JF, Bernard HR, Shields EP, Lindsey ML, Yazici A. 1992. A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgeries performed. *Health Services Research* **27**: 517–542.

Hannan EL, Siu AL, Kumar D, Kilburn Jr H, Chassin MR. 1995. The decline in coronary artery bypass graft surgery mortality in New York state. The Role of Surgeon volume. *Journal of the American Medical Association* **273**: 209–213.

Ho V. 2000. Evolution of the volume–outcome relation for hospitals performing coronary angioplasty. *Circulation* **101**: 1806.

Hughes RG, Garnick DW, Luft HS, McPhee SJ, Hunt SS. 1988. Hospital volume and patient outcomes: the case of hip fracture patients. *Medical Care* **26**: 1057–1067.

Jha AK, Orav EJ, Epstein AM. 2007. The inverse relationship between mortality rates and performance in the hospital quality alliance measures. *Health Affairs* **26**: 1104–1110.

Luft HS, Bunker JP, Enthoven A. 1979. Should operations be regionalized? *The New England Journal of Medicine* **301**: 1364–1369.

Luft HS. 1980. The relation between surgical volume and mortality: an exploration of causal factors and alternative models. *Medical Care* **18**: 940–959.

Luft HS. 1998a. From observing the relationship between volume and outcome to making policy recommendations. *Medical Care* **41**: 1118–1122.

Luft HS. 1998b. Pushing good research to go farther. *Health Services Research* **33**: 1185–1190.

Luft HS. 2003. From observing the relationship between volume and outcome to making policy recommendations: Comments on Sheikh. *Medical Care* **41**(10): 1118–1122.

Luft HS, Garnick DW, Mark DH, Peltzman DJ, Phibbs CS, Lichtenberg E, McPhee SJ. 1990. Does quality influence choice of hospital? *Journal of the American Medical Association* **263**: 2899–2906.

Marcin JP, Li Z, Kravitz RL, Dai JJ, Rocke DM, Romano PS. 2008. The CABG surgery volume–outcome relationship: temporal trends and selection effects in California, 1998–2004. *Health Services Research* **43**: 174–192.

Nallamothu BK, Lu X, Vaughan-Sarrazin MS, Cram P. 2008. Coronary revascularization at specialty hospitals and peer general hospitals in black medicare beneficiaries. *Circulation Cardiovascular Quality and Outcomes* **1**: 116–122.

DOI: 10.1002/hec

Norton EC, Garfinkel SA, McQuay LJ, Heck DA, Wright JG, Dittus R, Lubitz RM. 1998. The effect of hospital volume on the in-hospital complication rate in knee replacement patients. *Health Services Research* **33**: 1191–1210.

Popescu I, Nallamothu BK, Vaughan-Sarrazin MS, Cram P. 2008. Do specialty cardiac hospitals have greater adherence to acute myocardial infarction and heart failure process measures? An empirical assessment using medicare quality measures: quality of care in cardia specialty hospitals. *American Heart Journal* **156**: 155–159.

Shahian DM, Normand SL. 2003. The volume–outcome relationship: from Luft to Leapfrog. *Annals of Thoracic Surgery* **75**: 1048–1058.

Sheikh K. 2003. Reliability of provider volume and outcome associations for healthcare policy. *Medical Care* **41**: 1111–1117.

Sowden AJ, Sheldon TA. 1998. Does volume really affect outcome? Lessons from the evidence. *Journal of Health Services Research and Policy* **3**: 187–190.

Stewart J, Gill L. 1991. *Econometrics*. London: Prentice Hall Europe.

Tsai AC, Votruba M, Bridges JF, Cebul RD. 2006. Overcoming bias in estimating the volume–outcome relationship. *Health Services Research* **41**: 252–264.

Vakili BA, Kaplan R, Brown DL. 2001. Volume-outcome relation for physicians and hospitals performing angioplasty for acute myocardial infarction in New York State. *Circulation* **104**: 2171–2176.

Vetrhus M, Sondenaa K. 2007. Hospital volume does not affect clinical outcome. *Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* **33**(9): 1049–1051.