Overcoming Bias in Estimating the Volume–Outcome Relationship

Alexander C. Tsai, Mark Votruba, John F. P. Bridges, and Randall D. Cebul

Objective. To examine the effect of hospital volume on 30-day mortality for patients with congestive heart failure (CHF) using administrative and clinical data in conventional regression and instrumental variables (IV) estimation models.

Data Sources. The primary data consisted of longitudinal information on comorbid conditions, vital signs, clinical status, and laboratory test results for 21,555 Medicare-insured patients aged 65 years and older hospitalized for CHF in northeast Ohio in 1991–1997.

Study Design. The patient was the primary unit of analysis. We fit a linear probability model to the data to assess the effects of hospital volume on patient mortality within 30 days of admission. Both administrative and clinical data elements were included for risk adjustment. Linear distances between patients and hospitals were used to construct the instrument, which was then used to assess the endogeneity of hospital volume.

Principal Findings. When only administrative data elements were included in the risk adjustment model, the estimated volume–outcome effect was statistically significant ($p = .029$) but small in magnitude. The estimate was markedly attenuated in magnitude and statistical significance when clinical data were added to the model as risk adjusters ($p = .39$). IV estimation shifted the estimate in a direction consistent with selective referral, but we were unable to reject the consistency of the linear probability estimates.

Conclusions. Use of only administrative data for volume–outcomes research may generate spurious findings. The IV analysis further suggests that conventional estimates of the volume–outcome relationship may be contaminated by selective referral effects. Taken together, our results suggest that efforts to concentrate hospital-based CHF care in high-volume hospitals may not reduce mortality among elderly patients.

Key Words. Volume-outcome effect, selective referral, instrumental variables, congestive heart failure, outcomes research

Twenty-five years ago (Luft, Bunker, and Enthoven 1979), documented an inverse volume–outcome relationship for surgical care. Subsequent studies in this literature (Halm, Lee, and Chassin 2000) have generated similar results. Accordingly, organizations such as the Leapfrog Group have encouraged

patients undergoing high-risk procedures to seek care at high-volume hospitals (Birkmeyer, Finlayson, and Birkmeyer 2001).

Congestive heart failure (CHF) is a particularly important area of investigation for improving quality of care, because of its high prevalence and considerable cost to the U.S. Medicare program, where the average charge for a CHF admission exceeds \$10,000 (U.S. Centers for Medicare and Medicaid Services 2003). Despite technological advances, 5-year survival rates for CHF have failed to improve. This is commonly attributed to the complexity of medical management of patients with CHF (Akosah et al. 2002), suggesting that a learning by doing effect should be observed in the care of such patients. However, few studies have estimated the volume–outcome relationship for nonprocedural care (Halm, Lee, and Chassin 2000). Only one previous study has been performed so in the context of CHF. Burns and Wholey (1991) analyzed 1,988 statewide administrative data on nearly 6,000 CHF admissions and concluded that there was no suggestion of an inverse relationship between hospital volume and in-hospital mortality.

Interpreting previous work on the inverse volume–outcome relationship, however, is subject to two important caveats, both of which are remedied in our study. The first caveat to the volume–outcome literature is that the estimated relationship may reflect unmeasured differences in the illness severity of patients admitted to low- versus high-volume hospitals. Most volume–outcome studies, including the one by Burns and Wholey (1991), have not used detailed clinical data for appropriate risk adjustment (Halm, Lee, and Chassin 2000) and may therefore be subject to unpredictable biases resulting from unmeasured severity of illness. Our analyses are based on clinical data that contain a much greater level of detail than that which is found in typical discharge abstract summaries. We then compare estimates based on administrative versus clinical risk adjustment models in order to illustrate the importance of appropriate risk adjustment.

Address correspondence to Alexander C. Tsai, Ph.D., Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine WG-57, 10900 Euclid Avenue, Cleveland, OH 44106-4945. Mark Votruba, Ph.D., Assistant Professor of Economics, is with the Department of Economics, Weatherhead School of Management, Case Western Reserve University, Cleveland, OH. John F. P. Bridges, Ph.D., Leader of the Junior Group, is with the International Health Economics and Outcomes Research, Department of Tropical Hygiene and Public Health, University of Heidelberg Medical School, Heidelberg, Germany. Randall D. Cebul, M.D., Director, is with the Center for Health Care Research and Policy at MetroHealth Medical Center, Cleveland, OH and also Professor of Medicine, Epidemiology, and Biostatistics at Case Western Reserve University, Cleveland, OH.

The second caveat is that the estimated relationship may reflect reverse causality, as providers known for high quality care will attract greater numbers of patients through “selective referral” (Luft 1980). Without the benefit of randomization, conventional estimates of the volume–outcome effect may be contaminated by selective referrals and will therefore exaggerate the importance of learning by doing. To identify the volume–outcome effect, previous studies have used the method of instrumental variables (IV), with the total number of hospital beds serving as an instrument (Luft, Hunt, and Maerki 1987; Hughes et al. 1988; Farley and Ozminkowski 1992; Norton et al. 1998). Using hospital bed size as an instrument, however, is not without its weaknesses. A legitimate instrument must also be conditionally unrelated with the dependent variable of interest except through its effect on the potentially endogenous regressor of interest. If patients are being selectively referred to generally higher-quality hospitals, then one would expect these hospitals to be larger in size, and CHF-specific quality of care is likely to be correlated with general quality of care provided by a given hospital. To remedy the shortcomings inherent in previous IV analyses, we construct an instrument that is based on the linear distances between patients and hospitals. This new instrument can serve as a plausibly exogenous source of variation in hospital CHF volume, which we use to appropriately identify the volume–outcome effect.

METHODS

Data Sources

The data used for this study consists of longitudinal information on demographic characteristics, vital signs, clinical status, comorbid conditions, and laboratory test results for Medicare-insured patients aged 65 years and older hospitalized for CHF in all nonfederal hospitals in northeast Ohio from 1991 to 1997. Full details regarding construction of the initial CHF patient sample have been published previously (Baker et al. 2002, 2003a). The initial sample consisted of 23,505 Medicare patients hospitalized with a first admission for CHF at CHQC hospitals. Patients with Medicare listed as their primary insurance comprised 79 percent of all CHQC patients hospitalized for CHF (Baker et al. 2003a). After applying additional exclusion criteria, we retained 21,841 observations in the analytic sample. The sample size was further reduced in the regression analyses because of missing data.

Using the U.S. Census Bureau Topologically Integrated Geographic Encoding and Referencing (TIGER) database, we matched each patient's zip code of residence to the latitude and longitude coordinates of the zip code centroid. Hospital addresses were matched to exact latitude and longitude coordinates. The Great Circle Distance formula was then used to compute the Euclidean distances between each latitude/longitude pair (Gowrisankaran and Town 1999; Kessler and McClellan 2000). We restricted the analysis to patients whose chosen hospital was within 75 miles of his or her residence (Gowrisankaran and Town 1999). This study was approved by the university Institutional Review Board.

Covariates

Death occurring within 30 days of admission was determined from the “Admission to Date of Death Interval” field in the MedPAR data files. For hospital volume, each patient admission was linked with the number of CHF patients admitted at that hospital in the calendar year prior to the year of the patient's visit. All patient observations in 1991 were therefore excluded from analysis, as hospital volumes prior to 1991 were unavailable. Hospital volume was specified as a continuous variable, scaled so that the estimated regression coefficients could be interpreted as the effect of a 100-patient increase in volume. To adjust for severity of illness on admission, the regression models included a set of administrative and clinical data elements that were used in a previously published CHF-specific risk adjustment model that had excellent discrimination and calibration, with a c -statistic of 0.84 and a χ^2 goodness-of-fit statistic of 10.9 ($p = .21$) (Baker et al. 2002, 2003a). Teaching status was defined according to the reported number of residents per bed set up and staffed for patient care (Taylor, Whellan, and Sloan 1999).

Least Squares Estimation

The patient was the unit of analysis. A linear probability regression model (LPM) was fit to the data to assess the effects of hospital volume on 30-day mortality:

$$MORT_{ijt} = \beta_0 + \beta_1 VOL_{j,t-1} + \beta_2 X_{ijt} + \varepsilon_{ijt} \quad (1)$$

where $MORT_{ijt}$ represents the outcome for patient i in hospital j in year t and is equal to unity if the patient died within 30 days of admission, zero otherwise; $VOL_{j,t-1}$ is the regressor of primary interest and represents the number of patients discharged from hospital j in year $t-1$ with a principal diagnosis of CHF; X_{ijt} is a vector of variables, including both administrative and clinical

adjusters for severity of illness, dummies for hospital teaching status, and dummies for year of admission; and the expected value of the error term ε_{ijt} is zero and uncorrelated with the primary regressor of interest, hospital volume. Variance-covariance estimators robust to heteroskedasticity (Huber 1967; White 1980) and clustering at the hospital-year level (Williams 2000) were used, and statistical significance tests were two-tailed throughout.

IV Estimation Using Expected Volume

Several studies have noted that distance is an important determinant of patients' choice of hospital (Luft et al. 1990; Gowrisankaran and Town 1999). Kessler and McClellan (2000) used a conditional logit model of hospital choice to construct measures of hospital market structure free of unobserved heterogeneity across individual patients, individual hospitals, and geographic hospital markets. We used their method to construct a measure of expected volume to serve as a source of identification for estimating the volume-outcome relationship.¹ For this portion of the analysis, we used an expanded sample of 43,469 Medicare and non-Medicare CHF admissions and readmissions. After excluding 1,046 observations representing patients who traveled more than 75 miles to their chosen hospital (Gowrisankaran and Town 1999) and another 41 observations from one hospital that only reported 2 years of data, we retained 42,382 CHF admissions for analysis. The patient's choice of one hospital over another was a function of the hospital's distance from the patient as well as its distance relative to other (similar) hospitals that the patient might regard as substitutes. Patient i 's indirect expected utility from choosing hospital j at time t (U_{ijt}^*) was modeled as follows:

$$\begin{aligned} U_{ijt}^* &= W_{ijt} + \omega_{ijt} \\ &= \delta_1 D_{ijt} + \delta_2 D_{ijt}^2 + \delta_3 D_{ijt}^3 + \delta_4 CLOSE_{ijt} + \delta_5 DIFF_{ijt} + \omega_{ijt} \end{aligned} \quad (2)$$

where D_{ijt} represents the Euclidean distance from the centroid of the patient's zip code to the hospital's address, with quadratic and cubic terms (each scaled for convenience by dividing by 100) included to capture potential nonlinear effects and $CLOSE_{ijt}$ is a dummy variable denoting whether the hospital is the closest hospital to the patient. The variable $DIFF_{ijt}$ represents the distance from patient i to the nearest similar hospital j' minus the distance from patient i to hospital j . Hospitals were defined as "similar" on the basis of bed size: hospitals were sorted into four categories of bed size, with the cutoff points roughly corresponding to the 25th, 50th, and 75th percentiles of bed size averaged over the entire study period: ≤ 190 beds, 190–249 beds, 250–330

beds, and >330 beds. Hospital j' was considered similar if it belonged to the same bed size category as hospital j . The model assumes that patients choose the hospital that maximizes the utility gained from that choice and that the error term ω_{ijt} is independently and identically distributed with a type I extreme value distribution.

The method of maximum likelihood was used to estimate the parameters of equation (2). The estimated parameters were used to compute the probability (π_{ijt}) of patient i choosing hospital j at time t . These probabilities were then aggregated over the sample to compute CHF volumes for each hospital that would be expected on the basis of the geographical distribution of hospitals and patients in each year (“expected volume”):

$$\hat{V}_{jt} = \sum_{i=1}^I \hat{\pi}_{ijt} = \sum_{i=1}^I \frac{\exp(\hat{U}_{ijt})}{\sum_{j=1}^J \exp(\hat{U}_{ijt})} \quad (3)$$

With this variable serving as an instrument, we used linear IV methods to consistently estimate β_1 in equation (1). Standard errors were robust to heteroskedasticity (Huber 1967; White 1980) and clustering at the hospital-year level (Williams 2000). The consistency of the linear probability estimates was assessed using the augmented regression test suggested by (Davidson and MacKinnon 1993). The F -test of the statistical significance of the instrument was used to ascertain that the F -statistic value was greater than 10 (Staiger and Stock 1997).

RESULTS

Table 1 provides descriptive statistics for the sample. In 1997, the average bed size across the hospitals in the sample was 276 (range 84–906), the average CHF case volume was 459 (77–1,203), and the average resident-to-bed ratio was 0.13 (0–0.69). Column 1 of Table 2 displays estimated regression coefficients for the linear probability regression model from equation (1) that included only administrative data elements as risk adjusters. Each 100-person increase in volume was negatively associated with a 0.2 percent absolute decline in 30-day mortality ($p = .029$), but the effect was small in magnitude. Column 2 of Table 2 displays estimated regression coefficients for the linear probability regression model that also included clinical data elements as risk adjusters. The proportion of variance in the outcome explained by the model

Table 1: Means of Variables, by Average Annual Hospital Volume Quartiles ($N = 21,841$)

	<i>Bottom Quartile[†]</i> (77–191)	<i>Middle Quartiles</i> (192–508)	<i>Top Quartile</i> (509–815)
Outcomes			
Died in hospital (%)	5.2	5.6	5.1
Died within 30 days (%)	9.8	9.6	9.3
Administrative data elements			
Age (years)	78.8***	77.9	77.8
Female (%)	59.1	58.6	55.5***
White (%)	92.2***	81.7	86.7***
Atrial fibrillation (%)	26.8	28.4	31.4***
Peripheral vascular disease (%)	8.6***	6.5	7.6**
Chronic obstructive pulmonary disease (%)	28.4**	25.2	23.1***
Currently on dialysis (%)	0.6***	2.0	2.0
Metastatic cancer (%)	1.6	1.5	1.5
Clinical data elements			
Pulse (beats per minute)	95.8	94.9	93.9***
Systolic blood pressure (mmHg)	149.8**	152.5	149.1***
Body temperature <96 or >100°F (%)	30.8***	38.3	32.0***
Coma, lethargy, or stupor (%)	3.1	3.5	3.3
Early do-not-resuscitate order (%)	14.3**	12.0	12.0
Blood urea nitrogen (mg/dL)	28.2*	29.3	30.2***
Serum sodium (mEq/L)	138.2	138.2	138.1
Serum albumin (g/dL)	3.6*	3.5	3.7***
Serum glucose <85 or >125 mg/dL (%)	70.2	69.8	65.9***
SGOT >30 IU/L (%)	40.8*	43.9	44.4
PCO ₂ <25 or >55 mmHg (%)	6.0	6.7	6.7
A-aO ₂ gradient (mmHg)	94.9	93.4	80.3***
Number of observations	2,052	10,469	9,320

[†]Hospitals were grouped into quartiles on the basis of average annual CHF case volume, where each hospital's average annual CHF case volume was computed by averaging across all years in 1992 through 1997 for which the hospital reported data.

* $p < .05$;

** $p < .01$;

*** $p < .001$. Statistical significance according to the Wilcoxon rank-sum test for comparison with the middle quartiles.

CHF, congestive heart failure.

increased from 0.024 to 0.170, and an F -test suggests that the clinical variables jointly affect the outcome ($F = 172$, $p < .001$). Notably, the statistical significance of the estimated regression coefficient on volume was markedly attenuated ($p = .39$).

The estimated coefficients from the hospital choice model are available upon request. As expected, a given hospital's distance from a patient was

Table 2: Effect of Hospital Volume on 30-day Mortality ($N = 21,439$)

	LPM		IV	
	(1) Administrative Data	(2) Clinical Data	(3) Administrative Data	(4) Clinical Data
Hospital volume	−0.002 [−2.20]	−0.0009 [−0.86]	−0.0001 [−0.05]	0.003 [1.24]
Minor teaching hospital	0.002 [0.30]	0.002 [0.36]	−0.001 [−0.15]	−0.003 [−0.40]
Major teaching hospital	0.008 [1.57]	−0.001 [−0.13]	0.004 [0.54]	−0.009 [−1.12]
1993*	−0.007 [−0.81]	0.004 [0.47]	0.001 [0.10]	0.018 [1.51]
1994	−0.008 [−0.90]	−0.0001 [−0.01]	−0.001 [−0.12]	0.012 [1.00]
1995	−0.016 [−2.01]	−0.003 [−0.38]	−0.011 [−1.09]	0.007 [0.67]
1996	−0.023 [−3.00]	−0.014 [−1.86]	−0.017 [−1.76]	−0.003 [−0.33]
1997	−0.022 [−2.77]	−0.007 [−0.13]	−0.017 [−1.71]	0.002 [0.23]
Intercept	−0.187 [−8.61]	0.207 [2.95]	−0.202 [−8.45]	0.183 [2.59]

Notes: t - and z -statistics are displayed in brackets beside the estimated coefficients. In addition to the variables reported in the table, the administrative risk adjustment models represented by columns 1 and 3 included the following variables: age, female sex, race (white), atrial fibrillation comorbidity, peripheral vascular disease, chronic obstructive lung disease, currently on dialysis, and metastatic cancer. The clinical risk adjustment models represented by columns 2 and 4 included the following variables: age, female sex, race (white), pulse, systolic blood pressure, temperature, coma or lethargy, do-not-resuscitate order written within 2 days of admission, atrial fibrillation comorbidity, peripheral vascular disease, chronic obstructive lung disease (directly abstracted from patient chart), currently on dialysis, metastatic cancer, blood urea nitrogen, serum sodium, serum albumin, serum glucose < 85 or > 125 , serum glutamic-oxaloacetic transaminase > 30 , partial pressure of carbon dioxide on ABG < 25 or > 55 , and arterial-alveolar oxygen gradient on ABG. The estimated coefficients on these variables are available on request from the authors.

*The dummy variable for 1992 was omitted as the reference category.

LPM, linear probability model; IV, instrumental variable; ABG, arterial blood gas.

inversely associated with the probability of choosing that hospital ($p < .001$, all years). The estimated coefficients on the quadratic and cubic terms were also highly statistically significant. Hospitals nearest in geographical proximity to the patient were more likely to be patronized ($p < .001$, all years). Collectively, the regressors explained approximately 40 percent of the variation in hospital choice.

The results of the linear IV analysis are displayed in columns 3 and 4 of Table 2. Column 3 provides the IV estimate when only administrative data elements were considered. Instrument relevance was high ($F = 75.5$, $p < .001$). Compared with the linear probability estimate in column 1, the IV estimate reversed sign but was statistically indistinguishable from zero ($p = .96$). While the IV estimate was shifted in a direction consistent with selective referral, the augmented regression test could not reject the consistency of the linear

probability estimate ($F = 1.19$, $p = .28$). Column 4 displays the IV estimate when clinical data were added to the model as risk adjusters. The IV estimate of the volume–outcome effect was positive but of marginal statistical significance ($p = .22$). Instrument relevance was high ($F = 75.1$, $p < .001$). The augmented regression test nearly rejected the consistency of the linear probability estimate ($F = 3.75$, $p = .055$). In all models, estimates of the volume–outcome effect were extremely small in magnitude.

We conducted several sensitivity analyses to determine whether our findings were robust to alternative specifications. First, to adjust for potentially relevant socioeconomic factors related to geographic location, we included four zip code-level variables derived from 1990 U.S. Census data: unemployment rate, percentage of residents living below the poverty level, percentage of black residents, and percentage of residents who did not finish high school. Second, to assess whether our findings were being driven by patients traveling longer distances, we recomputed the expected volume measure by refitting equation (2) to a dataset limited to the 32,537 patients living in the four counties represented by the CHQC hospitals. This new instrument, which was closely correlated with the original instrument, was then used to obtain revised IV estimates. Third, we also computed alternative expected volume IVs based on different combinations of the variables in equation (2). Fourth, we tried different specifications for hospital volume, including log volume, dummy variables with different cutoffs, and volume in the 365 days before the day of admission. Fifth, we added “distance to nearest hospital” to the volume–outcome regression models to proxy for how far a patient is likely to have to travel for care. Sixth, we fit probit and nonlinear (probit) IV models (Newey 1987) corresponding to the linear IV models. All of these sensitivity analyses generated qualitatively similar point estimates on hospital volume, reinforcing our confidence in our original results.

DISCUSSION

Two competing hypotheses relevant to policy making in this area are the “learning by doing” and “selective referral” hypotheses. Appropriate policy making requires disentangling these two effects, because the policy implications will differ depending on which of the two, on balance, is more likely to be true. If the learning by doing hypothesis is true, then diverting patients from low-volume hospitals to high-volume hospitals using a strategy of volume-based referral (Birkmeyer, Finlayson, and Birkmeyer 2001) should lead to

improved outcomes for patients who switch to high volume hospitals (but not for the remaining patients who are still being treated at low-volume hospitals).

If the selective referral hypothesis is true, then conventional estimates of the volume–outcome effect will tend to be exaggerated. To account for the potential endogeneity of hospital volume, we used the method of IV. In contrast to earlier IV analyses in the volume–outcome literature, the instrument used in our paper—the volume of CHF patients that would be expected at a hospital given the geographical distribution of patients and hospitals, or “expected volume”—was an appealing choice because it was based on observable, exogenous factors related to the geographical distribution of patients and hospitals. Although the administrative risk adjustment model suggested an inverse volume–outcome relationship, the effect was small in magnitude. We speculate that, if an inverse volume–outcome relationship truly exists for the care of patients with CHF, it may be more salient to study the effect of physician-specific volume. Applying the method of IV shifted the estimated volume–outcome effects in a direction consistent with the selective referral hypothesis, suggesting that hospital volume is endogenously determined.² We were unable to reject the consistency of conventional estimates, but this may be attributable to inadequate sample size and imprecision of the IV estimates.

Our study also emphasizes the importance of using clinical data for risk adjustment. One recent review of 88 volume–outcome studies noted that the majority of these studies relied solely on administrative data elements such as ICD-9 discharge diagnoses (Halm, Lee, and Chassin 2000). Only four studies reported robustly discriminating risk adjustment models. In our study, when additional clinical variables were included—including information on vital signs, clinical status, and laboratory test results—the volume–outcome estimate was substantially attenuated in magnitude and statistical significance. This result is consistent with a systematic review that documented more anticonservative findings in studies that used more primitive risk adjustment methods (Sowden, Deeks, and Sheldon 1995).

CONCLUSION

This paper advances the volume–outcome literature by illustrating the consequences of poor risk adjustment and by accounting for the potential endogeneity of hospital volume with an instrument that can be easily constructed for use in other clinical contexts. Our administrative risk adjustment model generated findings supportive of the “learning by doing” hypothesis that were

found to be spurious when clinical data elements were added as risk adjusters. We also note that using the method of IV to account for potential endogeneity yielded estimates that were consistent with the “selective referral” hypothesis. Taken together, these results suggest that efforts to concentrate hospital-based CHF care in high-volume hospitals may not reduce mortality. Policies such as volume-based referral have been recommended as a way to improve surgical care, but doing so for CHF care may not be indicated until more definitive evidence has accumulated.

ACKNOWLEDGMENTS

At the time this research was conducted, Dr. Tsai was a National Research Service Award Trainee supported by the U.S. Agency for Healthcare Research and Quality (AHRQ) Institutional Training Award T32 HS 00059-06). This work was also supported in part by the U.S. AHRQ Dissertation Research Grant 1 R36 HS 014151-01 and the Harvey Fellows Program of the Mustard Seed Foundation. Scott Husak provided invaluable data management assistance. We also appreciate helpful comments from Avi Dor, Jim Rebitzer, J. B. Silvers, an anonymous reviewer, and seminar participants from the Center for Health Care Research and Policy at MetroHealth Medical Center.

NOTES

1. This concept is similar to that of Chernew, Gowrisankaran, and Fendrick (2002), although they estimate a model of patient flows for coronary artery bypass graft surgery to use in an entry model. We thank an anonymous reviewer for bringing this reference to our attention.
2. This finding is consistent with a previously published study by Baker et al. (2003b) that concluded, based on the same CHQC data, that public reporting of hospital quality in this region did not affect patient flows.

REFERENCES

- Akosah, K. O., A. M. Schaper, P. Havlik, S. Barnhart, and S. Devine. 2002. “Improving Care for Patients with Chronic Heart Failure in the Community: The Importance of a Disease Management Program.” *Chest* 122 (3): 906–12.

- Baker, D. W., D. Einstadter, C. Thomas, and R. D. Cebul. 2003a. "Mortality Trends for 23,505 Medicare Patients Hospitalized with Heart Failure in Northeast Ohio." *American Heart Journal* 146 (2): 258–64.
- Baker, D. W., D. Einstadter, C. Thomas, S. Husak, N. H. Gordon, and R. D. Cebul. 2003b. "The Effect of Publicly Reporting Hospital Performance on Market Share and Risk-Adjusted Mortality at High-Mortality Hospitals." *Medical Care* 41 (6): 729–40.
- Baker, D. W., D. Einstadter, C. L. Thomas, S. S. Husak, N. H. Gordon, and R. D. Cebul. 2002. "Mortality Trends during a Program That Publicly Reported Hospital Performance." *Medical Care* 40 (10): 879–90.
- Birkmeyer, J. D., E. V. Finlayson, and C. M. Birkmeyer. 2001. "Volume Standards for High-Risk Surgical Procedures: Potential Benefits of the Leapfrog Initiative." *Surgery* 130 (3): 415–22.
- Burns, L. R., and D. R. Wholey. 1991. "The Effects of Patient, Hospital, and Physician Characteristics on Length of Stay and Mortality." *Medical Care* 29 (3): 251–71.
- Chernew, M., G. Gowrisankaran, and A. M. Fendrick. 2002. "Payer Type and the Returns to Bypass Surgery: Evidence from Hospital Entry Behavior." *Journal of Health Economics* 21 (3): 451–74.
- Davidson, R., and J. G. MacKinnon. 1993. "Instrumental Variables." In *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Farley, D. E., and R. J. Ozminkowski. 1992. "Volume–Outcome Relationships and In-Hospital Mortality: The Effect of Changes in Volume over Time." *Medical Care* 30 (1): 77–94.
- Gowrisankaran, G., and R. J. Town. 1999. "Estimating the Quality of Care in Hospitals Using Instrumental Variables." *Journal of Health Economics* 18 (6): 747–67.
- Halm, E. A., C. Lee, and M. R. Chassin. 2000. "How Is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature." In *Interpreting the Volume–Outcome Relationship in the Context of Health Care: Workshop Summary*. Washington, DC: Institute of Medicine.
- Huber, P. J. 1967. "The Behavior of Maximum Likelihood Estimates under Non-standard Conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.
- Hughes, R. G., D. W. Garnick, H. S. Luft, S. J. McPhee, and S. S. Hunt. 1988. "Hospital Volume and Patient Outcomes. The Case of Hip Fracture Patients." *Medical Care* 26 (11): 1057–67.
- Kessler, D. P., and M. B. McClellan. 2000. "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics* 115 (2): 577–615.
- Luft, H. S. 1980. "The Relation between Surgical Volume and Mortality: An Exploration of Causal Factors and Alternative Models." *Medical Care* 18 (9): 940–59.
- Luft, H. S., J. P. Bunker, and A. C. Enthoven. 1979. "Should Operations Be Regionalized? The Empirical Relation between Surgical Volume and Mortality." *New England Journal of Medicine* 301 (25): 1364–9.
- Luft, H. S., D. W. Garnick, D. H. Mark, D. J. Peltzman, C. S. Phibbs, E. Lichtenberg, and S. J. McPhee. 1990. "Does Quality Influence Choice of Hospital?" *Journal of the American Medical Association* 263 (21): 2899–906.

- Luft, H. S., S. S. Hunt, and S. C. Maerki. 1987. "The Volume–Outcome Relationship: Practice-Makes-Perfect or Selective-Referral Patterns?" *Health Services Research* 22 (2): 157–82.
- Newey, W. K. 1987. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics* 36 (3): 231–50.
- Norton, E. C., S. A. Garfinkel, L. J. McQuay, D. A. Heck, J. G. Wright, R. Dittus, and R. M. Lubitz. 1998. "The Effect of Hospital Volume on the In-Hospital Complication Rate in Knee Replacement Patients." *Health Services Research* 33 (5, part 1): 1191–210.
- Sowden, A. J., J. J. Deeks, and T. A. Sheldon. 1995. "Volume and Outcome in Coronary Artery Bypass Graft Surgery: True Association or Artefact?" *British Medical Journal* 311 (6998): 151–5.
- Staiger, D. O., and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586.
- Taylor, D. H. Jr., D. J. Whellan, and F. A. Sloan. 1999. "Effects of Admission to a Teaching Hospital on the Cost and Quality of Care for Medicare Beneficiaries." *New England Journal of Medicine* 340 (4): 293–299.
- U.S. Centers for Medicare and Medicaid Services. 2003. *2001 Medicare and Medicaid Statistical Supplement to the Health Care Financing Review*. Baltimore: Centers for Medicare and Medicaid Services.
- White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 4 (4): 817–38.
- Williams, R. L. 2000. "A Note on Robust Variance Estimation for Cluster-Correlated Data." *Biometrics* 56 (2): 645–6.