Hospital Volume and Quality of Care

Principal Investigators:
Gabriel Picone (University of South Florida)
James Jollis (Duke University).

Other team members:
Justin Trogdon (University of Adelaine)
Martin Salm (Duke University).

Organization:
University of South Florida
4202 E. Fowler Ave., FAO 126
Tampa, FL 33620

Project Period: 9/1/2003 – 8/31/2005
Federal Project Officer: Marge Keyes

This research was supported by a grant from the Agency for Healthcare Research and Quality, "Hospital Volume and Quality of Care" (1R03-HS013386-01A1).

1. Structured Abstract

a. *Purpose*: This project has two aims: 1) obtain unbiased estimates of the effect of hospital volume on health outcomes and 2) analyze potential explanations for this causal relationship.

b. *Scope*: Hospitals that treat a large number of patients will gain experience about the most effective treatment methods. One would expect this expertise to be reflected in better health outcomes for these patients (practice makes perfect). However, using observational data, a selection bias arises, because patients prefer the perceived "better-quality" hospital (selective referral).

c. *Methods*: We used data from the Cooperative Cardiovascular Project (CCP) merged with the American Hospital Association's (AHA) annual survey of hospitals and the National Inpatient Survey (NIS). Our study combines detailed medical chart data with a differential distance instrumental variable approach to account for the endogeneity of volume. The dependent variables were 30-day mortality, 1-year mortality, and 30-day readmission. We conducted separate analysis for urban and rural areas.

d. *Results*: The results indicate that hospital volume is positively correlated with all our measures of quality of care for urban patients. The effect of hospital volume decreases substantially after we control for hospital characteristics. Controlling for endogenous selective referrals does not affect the coefficients of hospital volume in the mortality equations. For rural patients, we found a much weaker relationship between volume and outcomes. We also found in a second paper that the results are robust to the econometric strategy and that, when using standard Medicare claims data, there is no magic method.

e. *Key Words*: Volume outcome, practice makes perfect, selective referrals.

2. Purpose

Physicians in hospitals that treat a large number of patients with a similar condition will gain experience about the most effective treatments methods. One would assume that this expertise should be reflected in better health outcomes (practice-makes-perfect hypothesis) for patients admitted to high-volume hospitals. However, using observational data, a selection bias arises, because patients prefer the perceived "better-quality" hospital, creating a spurious correlation between volume and outcomes (selective-referrals hypothesis).

Although the positive relationship between hospital volume and quality of care has been well documented in the literature for a large number of medical interventions, few studies have tried to control for the selectivity bias due to selective referrals. Two approaches have been used to account for selective referrals. One approach is to use an instrumental variables methods, and the other is to use longitudinal data with hospital fixed effects. The instrumental variables methods depend crucially on the quality of the instruments, and studies using longitudinal data with hospital fixed effects have the drawback that hospital quality changes considerably through time.

Knowing if there is a causal relationship between hospital volume and health outcomes has important policy implications for several reasons. First, federal and state regulators faced a considerable number of decisions involving mergers and entry into markets when a certificate of need is required. Market advocates oppose mergers and certificate of need legislation, because they believe that competition results in lower cost and enhanced quality as opposed to regulated markets. However, if the practice-makes-perfect hypothesis is correct, then free markets may lead to less-than-optimal utilization of some procedures in hospitals, causing a decrease in quality. Second, insurers like Medicare and Medicaid would find it beneficial in terms of mortality and cost to encourage selective contracting. Third, for many patients, survival may improve by directing them to a high-volume hospital.

The original proposal had two specific aims:

1) *Obtain unbiased estimates of the effect of hospital volume on health outcomes*. This aim led to two papers. We attached both papers to this report.
    a. Paper 1: *Hospital Volume and Quality of Care: Selective referrals or Practice Makes Perfect?* This is the main paper of the project, in which we empirically analyzed whether the relation between hospital volume and outcomes is due to selective referrals or practice makes perfect. This paper was submitted to an economic journal, and we are waiting for its decision.
    b. Paper 2: *Estimating Treatment Effects from Medicare Claims Data*. This is a methodological paper that complements the first paper by testing the robustness of the econometric strategy and instruments. We estimate the basic empirical specification from paper 1 using different estimators commonly used in the treatment literature with a different set of controls. The estimators considered include regression methods, propensity to score methods, matching estimators, instrumental variable methods, and OLS. The goal is to analyze whether some estimators work better to account for the omitted variable bias of standard Medicare claims data.

This paper is currently being polished, and it will be submitted to an economic journal before the end of December.

2) *Analyze potential explanations for the causal relationship*: We do not have a paper on this aim yet, but we are currently working on the empirical analysis on this aim. We are currently empirically analyzing whether hospital volume is correlated with the probability of a patient receiving CABG, PTCA, or neither---conditional on surviving the original heart attack. Our goal is to understand what may cause the practice-make-perfect effect. We anticipate this paper to be finish during Spring 2006.

3. Scope

3.1 Background: There is a large amount of literature establishing that patients who are treated by hospitals with high volumes have better clinical outcomes than patients who were treated by hospitals with low volumes. Although the positive relationship between quality and hospital volume has been strongly established in the medical literature, it is not clear if this relationship is causal or spurious. There are two reasons why there may be a causal relationship between hospital volume and hospital quality. The first reason is that experience accumulated over time by treating certain types of patients makes hospitals better (practice-makes-perfect concept) at treating the patients, leading to better outcomes. This is the experience, or "learning," effect. The second reason is related to economies of scale; hospitals at certain level of output achieve optimal outcomes. This scale effect may be because hospitals that treat a large number of similar patients or performed a large number of similar procedures can purchase better and more expensive equipment compared with low-volume hospitals. Higher-volume hospitals also may have access to better nursing staff and equipment to deal with rare complications.

However, the positive relationship between volume and quality of care may be spurious if some unmeasured factors account for this relationship. For example, if better hospitals attract more patients, because their quality is correctly perceived to be higher (selective referral), then we will find a positive correlation between volume and outcomes. In this case, there will be reverse causation: better outcomes cause higher volume. Another form of spurious correlation would occur if some hospitals achieve a high volume by admitting a large number of healthy patients for some procedures, creating a fictitious relationship between the procedure and outcomes. Most studies that found this positive relationship between volume and quality of care have relied on cross-sectional comparison and have not attempted to determine if there is a causal relationship.

Heart disease is the leading cause of mortality in the United States, and most of these deaths involve AMIs. Because of the sudden nature of heart attacks, doctors must make life or death decisions very rapidly. AMI is a good candidate to have a strong correlation between experience and the quality of the outcomes. Not surprisingly, several studies have found a strong positive correlation between AMI volume and health outcomes. For example, we found that 30-day mortality for patients admitted to hospitals in the lowest quartile of AMI admissions was 12% higher than for patients admitted to hospitals in the highest quartile. The sudden nature of heart attacks also makes the "selective-referral" bias less likely because, in the case of AMIs, more rapid treatment, specifically aimed at opening the occluded artery, limits damage and lower mortality.

Thus, the emergency medical service generally is directed toward the nearest hospital. However, hospitals are often clustered so closely together that transportation delays for some patients are insignificant if they go to a higher-quality hospital. These marginal patients are likely to be referred to the higher-quality hospital.

4. Methods

4.1 Data

The primary source of data is the Cooperative Cardiovascular Project (CCP). The CCP sample consists of randomly selected patient records for Medicare patients admitted to nonfederal acute-care hospitals in the US with a primary diagnosis of acute myocardial infarction (AMI). All the sampling occurred between February 1994 and July 1995. The sample includes all hospitals in the US that had not participated in a four-state pilot phase (Alabama, Connecticut, Iowa, and Wisconsin). Medical records for each sampled hospitalization were forwarded to clinical abstraction centers. Altogether, charts were abstracted for approximately 180,000 AMI patients. In this study, we restricted our sample to patients who were admitted from home. We also excluded patients who were admitted to hospitals that had fewer than five AMI admissions the previous year. The data are linked with Medicare claims data and with the national death index. Thus, we know how long the patient survived after the index admission and if he/she was readmitted to the hospital.

We complemented the CCP data set with hospital variables from two sources. First, we obtained information on hospital ownership; cardiac intensive care beds; total number of beds; availability of facilities for cardiac catheterization, angioplasty, and open-heart surgery at the hospital; and the ratio of full-time residents to beds from the American Hospital Association's (AHA) annual survey of hospitals for different years. Second, we calculated hospital volume from the National Inpatient Sample (NIS), which is part of the Healthcare Cost and Utilization Project (HCUP). NIS contains all discharge data from approximately 1,000 hospitals located in 22 states between 1988 and 1998. NIS includes primary and secondary diagnoses and procedures, which allowed us to measure the number of AMI patients that a hospital admitted each year. Using NIS to measure hospital volume, rather than measuring hospital volume directly from CCP, has two advantages: 1) it includes all AMI patients admitted to the hospitals, rather than only Medicare fee-for-service patients (the median hospital volume using all admissions is over 200 admissions per year; if we were going to only use Medicare volume, the median would have been below 100 admissions per year), and 2) it allows us to measure hospital volume for the year before the patient was admitted.

Because not all hospitals from CCP are included in NIS, our preferred AMI hospital volume measure was missing for many hospitals. We used multiple imputation methods to "fill in" the missing volume information using the following strategy. First, we sampled with replacement from the set of hospitals with both NIS and CCP volume information. Then, we regressed the NIS volume on the Medicare hospital volume as measured in the CCP, a metropolitan statistical area dummy, the percentage of the population in the county over 65 and interaction terms among the regressors. The R-squared was over 60 percent for the imputation equation. Next, we used these coefficients to predict the total AMI hospital volume for the full sample of hospitals and defined our low- and high-volume cutoffs. This process was repeated 15 times to generate 15 data sets with imputed AMI hospital volume for all hospitals.

Finally, we ran the empirical models for each of the imputed data sets and combined the estimates using standard multiple imputation aggregation formulas.

We conducted separate analyses for patients who lived in metropolitan statistical areas (77,824 admissions) and patients who lived in non-metropolitan statistical areas (22,193 admissions). Hereafter, we denote a metropolitan statistical area as an "urban" area and a nonmetropolitan statistical area as a "rural" area.

4.2 Study Design

4.2.1 Paper 1: Hospital Volume and Quality of Care: Selective Referral or Practice Makes Perfect?

Dependent Variables: The main dependent variables were binaries, indicating mortality at 1 month and 1 year following AMI. We also defined a binary variable, indicating readmission for ischemic heart disease at 1 month following the date of the hospital discharge for the original AMI admission.

Hospital Volume: The key explanatory variable is hospital volume. As explained above, we used a categorical specification for hospital volume. We defined a high-volume hospital as one in the top quartile of admissions per year and a medium-volume hospital as one in the second and third quartiles of admissions per year. The hospital volume dummies were calculated separately for urban and rural areas in each imputed sample. For an urban area, a hospital is considered high volume if it has over 364 admissions per year on average across imputed samples (sd=6.8); it is considered a medium volume hospital if it has between 169 (sd=3.2) and 363 admissions per year. For rural hospitals, the thresholds are 152 for high volume and 93 to 151 for medium volume.

Sociodemographic characteristics: Sociodemographic variables included age at the time of the admission, gender, and race.

Hospital characteristics: We also controlled for the following hospital characteristics that may affect hospital quality and outcomes: the number of beds; the number cardiac intensive care beds; and binary variables for catheterization lab, cardiac surgery, teaching status (non-teaching hospitals do not have medical residents, minor teaching hospitals have resident-to-bed ratios below the national median, and major teaching hospitals have resident-to-bed ratios above the national median), and ownership status (government, for profit, or nonprofit).

Severity of illness and coexisting conditions: An advantage of CCP compared to Medicare claims data and standard surveys is that it contains very detailed information on patients' coexisting conditions (comorbidities) and severity of illness. We summarized all coexisting conditions by the Charlson index with binaries for lack of mobility and incontinence. This was a weighted sum of comorbidities, for which the weights are proportional to the risk of death from each comorbidity. Higher values of the Charlson index indicated worse health. Lack of mobility and incontinence were among the most important predictors of 1-year mortality.

Severity of illness conditions included log mean arterial pressure, pulse, time since the start of chest pains to admission, anterior myocardial infarction, ST elevation on admission electrocardiogram, Killip class, left ventricular ejection fraction, congestive heart failure, cardiac arrest, shock on arrival, hemorrhage, and renal function. All these variables were shown to be important determinants of health outcomes for AMI patients.

Killip class, a measure of severity at admission, is used as the initial assessment of the severity of the heart attack. Heart attack patients are classified into one of four

classes: those with no evidence of congestive heart failure (CHF) (1), those with mild to moderate CHF (2), those with overt pulmonary edema (3), and those in cardiogenic shock (4). Thus, a higher classification indicates a more severe heart attack. Left ventricular ejection fraction (LVEF) measures the relative output of blood per contraction and is grouped into three categories: <20%, 20% to 39%, and >39%. Low values of LVEF are associated with higher mortality. Renal dysfunction is also associated with higher mortality and is defined as creatinine >2.5 mg/dL or blood urea nitrogen >40 mg/dL.
Instruments: We included the differential distance between the nearest hospital and the nearest medium-volume hospital as well as between the nearest hospital and the nearest high-volume hospital. We calculated the distance from patients' residences to hospitals as the linear distance from the center of the patient's five-digit zip code to the center of the hospital's five-digit zip code. We used the same thresholds (250 and 500 admissions per year) to define medium- and high-volume hospitals for urban and rural patients. Finally, we also included the average hospital volume for all hospitals in a 50-mile radius of the patient's residence as an additional instrument.
Econometrics Strategy: We estimated the model using a maximum likelihood estimator, accounting for selective referrals. The paper named below discusses in more detail the likelihood function and the implementation.

4.2.2 Paper 2: Estimating Treatment Effects from Medicare Claims Data

The empirical specification used in paper 2 is identical to the empirical specification of paper 1, which was described in 4.2.1, with the following exceptions. First, we only used data for urban patients. Second, we defined a high-volume hospital using a binary variable, with a threshold of 169 admissions per year. Third, we analyzed the effect of catheterization and admission to a for-profit hospital on mortality. In this analysis, our main explanatory variables were always binary variables, and the effect of each binary variable on the dependent variable was known as the average treatment effect.

Traditionally, health economists have relied on instrumental variables methods to estimate average treatment effects, but, in recent years, a number of alternative estimators, such as propensity score and matching estimators, have been used to study these problems. In economics, these alternative estimators have been used mainly in the social program evaluation literature, though the origin for most of these estimators was in the medical treatment literature. Each of these estimators makes a different set of assumptions and may even estimate different effects, which has created a great deal of confusion about the best econometric strategy.

The goal of this paper is to compare the performance of different estimators commonly used in the treatment literature to estimate average treatment effects on 30-days and 1-year mortality for AMI patients. The estimators analyzed are 1) OLS, 2) probit, 3) regression methods, 4) propensity score, 5) matching, 6) 2SLS, 7) maximum likelihood (MLE), and 8) Abadie's causal IV estimator.

The strategy of this paper was to use our rich data set to analyze how sensitive these different estimators are to the addition of detailed severity-of-illness measures. We divided our covariates into three groups: (i) only demographic information about individuals; (ii) demographic information and detailed measures of severity of illness; and (iii) demographic information, measures of severity of illness, and measures of hospital characteristics. Our assumption is that the estimates obtained from specification (iii), or "methods with controls" (OLS, probit, regression methods, propensity score,

matching), using the regressors that included detailed measures of severity and hospital characteristics would be unbiased and efficient. Thus, we considered those estimates as the benchmark and compared these to estimates obtained using "methods as instruments" (2SLS, MLE, Abadie's causal IV estimator) with the regressors in specification (i) to evaluate the capability of instrumental variable methods in controlling for unobserved heterogeneity.

4.2.3 Paper 3: Exploring the causes behind learning by doing using CCP data

The dependent variable in this paper was whether a patient received PTCA, CABG, or neither after a heart attack. We restricted our sample to patients who survived at least 7 days after the AMI. We estimated the model using a multinomial probit model and allowed for the disturbances to be correlated.

Our main explanatory variables were the hospital volume binaries used in paper 1 and defined in 4.2.1. We also controlled for sociodemographic characteristics, hospital variables, and the severity-of-illness controls used in paper 1 (see 4.2.1). We interacted the hospital volume binaries and different severity-of-illness measures to test whether high-volume hospitals treated patients with similar conditions differently from low-volume hospitals. Our goal was to test whether high-volume hospitals achieved better outcomes by treating patients more aggressively than low-volume hospitals did.

5. Results

Paper 1: Without controlling for other factors, medium-volume hospitals (25th to 75th percentile of AMI admissions per year) and high-volume hospitals (over 75th percentile of AMI admissions per year) have lower mortality and readmission rates compared to low-volume hospitals for urban patients. For rural patients, high-volume hospitals have lower mortality rates than low-volume hospitals, but medium-volume hospitals have higher readmission rates than low-volume hospitals (significant at the ten percent level for medium-volume hospitals). Adding sociodemographic characteristics does not reduce the significance or magnitude of the coefficients for urban patients, but the effect of high-volume hospitals on 1-year mortality for rural patients is no longer significant. Given the average values of the dependent variables, admission to a high-volume hospital for an urban patient reduces 30-day mortality by 12.7%, 1-year mortality by 9.9%, and 30-day readmissions by 8.8%. These are important effects; if causal, policies that redirect patients to high-volume hospitals would have significant benefits on health outcomes.

Adding hospital characteristics reduces considerably the average treatment effects for all our measures of health outcomes. For example, for urban patients, the effect of an admission to a high-AMI-volume hospital on 30-day mortality is reduced from –0.022 to –0.015; for 1-year mortality, the effect goes from –0.030 to –0.019; for 30-day readmission, the effect goes from –0.012 to –0.004. This latter effect is no longer significant at conventional statistical levels. Controlling for coexisting conditions and severity of illness for urban patients does not affect the average treatment effects, but the significance of the associated coefficients is lower. Admission to a high-volume hospital reduces 30-day mortality by 5.8%, 1-year mortality by 6.3%, and 30-day readmissions by 4.4% compare to low-volume hospitals. Thus, the potential benefits due to better health outcomes from redirecting patients away from low-volume hospitals appears to be modest.

For rural patients, controlling for hospital characteristics, severity of illness, and coexisting conditions makes the estimates imprecise and not significant. For these patients, it is probably more important to have a hospital with decent AMI facilities within a manageable distance than to be triaged to a high-AMI-volume hospital.

The Rivers and Vuong test for endogeneity was negative for all the specifications, suggesting that that v_1 and v_2 are negatively correlated, which is consistent with selective referrals. However, the tests are only significant for 30-day and 1-year mortality for rural patients. We found no strong evidence of endogeneity and selective referrals for urban patients based on these tests.

Controlling for selective referrals produced similar results for the average treatment effect of the probit estimates after controlling for coexisting conditions and severity of illness for the mortality equations. The correlation coefficient was not significant in any of the mortality specifications, which implies that we cannot reject the null hypothesis that, after controlling for coexisting conditions and severity of illness, hospital volume is exogenous in the mortality equations (a result consistent with the Rivers and Vuong test). We did find evidence of endogeneity for 30-day readmission; the correlation coefficient was negative and significant, and the binaries describing hospital volume were positive and significant at the 10% level for high volume. This implies that patients admitted to a high-volume hospital have a higher probability of readmission. A plausible explanation for this result is that high-volume hospitals lower mortality but create, at the margin, patients more likely to be readmitted now that they have barely survived.

The most important determinants of 30-day mortality, 1-year mortality, and 30-day readmission for all patients are age (+ on mortality and – on readmissions) and many of the variables grouped as coexisting conditions and severity of illness. This group includes high Charlson index (+), lack of mobility (+), incontinence (+), log mean arterial pressure (–), number of hours since chest pain started (+), anterior myocardial infarction (+), higher Killip class (+), low left ventricular ejection fraction (+), and renal function problem (+). In contrast, very few hospital characteristics are significant at conventional levels (the main exception is major teaching, which has a negative effect on mortality and positive on re-admission). In many instances, variables have opposite signs in the mortality and readmission equations (e.g., age, major teaching, immobility, and incontinence). This is because the most severely ill patients are more likely to die and thus are less likely to be readmitted.

Three weaknesses of this study should be acknowledged. First, our sample was restricted to AMI admissions, and we did not study the effects of the specific treatment that the patient received. For example, after surviving a heart attack, a patient may have surgery or angioplasty to open the blocked arteries. It is possible that the practice-makes-perfect hypothesis is more important when performing procedures like surgery compared with treating patients who have similar conditions right after a health shock. Second, our imputed measure of hospital AMI volume introduces measurement error. The instrumental variable estimates could be correcting for measurement error as well as endogeneity, as evidenced by the small increase in the estimates when moving from the probit to the multivariate probit model. Finally, our sample contains only those individuals with AMI who survived long enough to be admitted to a hospital. Policies that close low-volume hospitals, which are often the closest hospital for a (potential) patient, would increase the number of individuals who die before reaching a hospital.

Thus, our results that find modest benefit from regionalization likely overstate the benefits to that type of policy.

Paper 2: Our results show that, for estimating average treatment effects, methods that rely on the selection on observables assumption (OLS, probit, regression, propensity score, and matching) produce almost identical results using similar sets of control variables and that there are no advantages over using a standard least squares regression when estimating the average treatment effect. For example, the average treatment effects of a hospital admission to a high-volume hospital on 1-year mortality using only demographics ranged from -0.013 (OLS, matching) to -0.014 (probit). Across methods and treatment studied, the addition of controls for severity of illness to the standard information contained in Medicare claims data has the largest impact on the estimated average treatment effects. In fact, conditional on detailed patient data, the addition of hospital characteristics has little impact on the average treatment effects. In the previous example, the average treatment effects ranged from -0.006 to -0.008 after we controlled for severity of illness; the range is between -0.004 and -0.006 when adding hospital characteristics. Methods that rely on differential distance as an instrument (2SLS, MLE, and Abadie's causal IV) tend to be very unstable across specifications with different sets of controls and have unreliable results. For example, for specification (iii), these estimates ranged from -0.009 (2SLS) to 0.025 (Abadie's IV).

This analysis has two important implications for researchers using hospital Medicare claims data or other types of hospital claims data. First, we found that differential distance did not perform well as an instrument in estimating average treatment effects for the applications that we studied. Second, we did not find evidence that estimation methods assuming selection on observables perform well in estimating average treatment effects in the absence of good data. In conclusion, authors should be careful when estimating causal effects using hospital claims data; the best situation is when we have detailed severity of illness and coexisting conditions for at least a subset of the data. Differential distance between the patient's residence and different types of hospitals, which at one time was considered an excellent instrument for the applications studied in this paper, should be used with caution, because it is not the magic bullet.

Paper 3: This will be the final paper on this project, and the main empirical analysis will be conducted after paper 2 is submitted to a journal (December 2005). Our work to date on this paper has been to create the relevant dependent and explanatory variables to be used in the analysis and some descriptive statistics. We have also worked on the econometric strategy and the conceptual framework. In our CCP sample, approximately 10% of patients underwent PTCA, and 5% of patients underwent CABG during the original hospital stay. These percentages differ significantly depending on whether the admitting hospital was a high-volume hospital or not.

6. List of Publications and Products

Picone G, J. Trogdon, and J. Jollis. *Hospital Volume and Quality of Care: Selective Referral or Practice Makes Perfect?* Unpublished Manuscript, University of South Florida, September 2005.

Picone G, A. Khwaja, J. Trogdon, et al. *Estimating Treatment Effects from Medicare Claims Data.* Unpublished Manuscript, University of South Florida, November 2005.