

Estimating the Quality of Care in Hospitals Using Instrumental Variables

Gautam Gowrisankaran

*Department of Economics
University of Minnesota*

and

Robert J. Town

*Graduate School of Management
University of California-Irvine*

JEL Classification Codes: I11, I12

Keywords: Mortality; Quality; Hospital; Instrumental Variables

This Version: June 15, 1999

Abstract

Mortality rates are a widely used measure of hospital quality. A central problem with this measure is selection bias: simply put, severely ill patients may choose high quality hospitals. We control for severity of illness with an instrumental variables (IV) framework using geographic location data. We use IV to examine the quality of pneumonia care in Southern California from 1989 to 1994. We find that the IV quality estimates are markedly different from traditional GLS estimates, and that IV reveals different determinants of quality. Econometric tests suggest that the IV model is appropriately specified, that the GLS model is inconsistent.

Corresponding author: Gautam Gowrisankaran, Department of Economics, University of Minnesota, 271 - 19th Avenue South, Minneapolis, MN 55455 USA. Telephone: 612-625-8310; Fax: 612-624-0209; E-mail: gautam@econ.umn.edu. We would like to thank Dan Akerberg, Lanier Benkard, Tom Buchmueller, David Cutler, Jon Gruber, Tom Holmes, Emmett Keeler, Yuichi Kitamura, Ariel Pakes, Maureen Smith, Frank Wolak and an anonymous referee for helpful comments. We also received helpful comments from seminar participants at the 1997 American Economics Association Winter Meetings in New Orleans, the University of Illinois, and the Eight Annual Health Economics Meetings in Minneapolis. Gowrisankaran acknowledges financial support from the Office of the Vice President for Research at the University of Minnesota.

1. Introduction

Measuring hospital quality is a vexing, albeit extremely important, problem facing health care researchers. Publicly available patient discharge databases allow researchers to easily calculate hospital-specific mortality rates which can serve as outcome-based measures of quality. However, the problem of selection bias complicates such a measure. Simply put, hospitals may differ in the severity of illness of the patients that they treat, as higher quality hospitals may attract a sicker patient population. Thus, the mortality rates for a hospital will have at least two components: one component reflects the severity of illness of the patients they treat and the other component reflects the quality of care they provide. In econometric terms, if a patient's choice of hospital is correlated with her (unobserved) severity of illness, then patient choice will be endogenous, and standard regression analysis will give inconsistent estimates of the hospital-specific contribution to mortality. For example, if high-quality hospitals attract sicker patients, those hospitals may not look very "good" after estimating their adjusted mortality rates.

Researchers have long recognized the selection bias problem and have devoted considerable effort to correcting for it. In most patient discharge databases there exists information (e.g. patient demographics, diagnoses, comorbidities) which can be used to formulate severity adjustment measures. Over the years, researchers have proposed many different adjustment mechanisms that make use of this patient-level information.¹ Unfortunately, the evidence presented here and elsewhere suggests that an important portion of severity remains unobserved even after controlling for observed patient characteristics.² If the unobserved component of severity of illness affects a patient's choice of hospital, then the selection problem remains and estimates of hospital quality will be incorrect.

An obvious solution to the selection problem is to gather additional clinical information from medical records and use this information to form a more accurate measure of severity of

¹ Iezzoni (1994) provides an excellent discussion of the methods that are available.

² See, for example, United States General Accounting Office (1994).

illness.³ However, it is enormously costly to collect clinical information. For instance, Keeler *et al.* (1990) and Keeler *et al.* (1992) spent over \$300 per patient-level observation and \$13,000 per hospital-level observation (approximately \$4 million in total) to control for severity of illness with clinical data. Therefore, it seems likely that clinical data will never be collected and disseminated to health care researchers in quantities necessary to supplant the use of discharge databases.

Thus, it is still important to be able to accurately estimate hospital quality using discharge databases. This is particularly true since hospital care is being delivered in a more “market” environment where there is an increasing desire among consumer groups, hospital regulators, HMO executives, hospital administrators, physicians, the media and the public-at-large for more information about the quality of care of hospitals. Policymakers from Congress to the county health boards can make more informed policy decisions that affect the hospital industry if they better understand how the hospitals under their jurisdiction are performing. While discharge databases from Medicare and many states can potentially be used to answer these questions, the tools to do so are still beset by the problem of selection bias.

The goal of this research is to develop and implement statistical techniques to correct for selection bias in hospital mortality figures. It is our hope that these adjusted mortality figures, obtained using the techniques suggested in this research, will more accurately represent the quality of care provided by hospitals. Our innovation is to control for selection bias from patient selection by using an instrumental variables (IV) method. As we detail below, we propose a simple theoretical model of hospital choice that allows us to use geographic location measures as instruments. IV methodology has a long history in economics and IV with geographic location measures as instruments has recently has been applied to measuring the technological effectiveness of different medical treatments.⁴

The IV methods that we use are designed to consistently estimate hospital quality by substituting “instruments” for the endogenous hospital choice variables in the mortality equation.

³ Again, Iezzoni (1994) provides an excellent review of the methods that are available to measure severity using medical record information.

⁴ See McClellan, McNeil, and Newhouse (1994) and McClellan and Newhouse (1997).

In order to produce consistent results, the instruments must be uncorrelated with unobserved severity of illness but correlated with a subset of the variables that predict hospital choice. To understand the validity of our instruments we briefly exposit our model of hospital choice. In our model, a patient's severity of illness is a function of her observed demographic characteristics and an unobserved component that is assumed to be identically distributed in the population.⁵ The patient and her physician make her choice of hospital based on many factors, including severity of illness, quality of care, and distance to the hospital.⁶ Since unobserved severity of illness is identically distributed in the population (by assumption), distance to a given hospital will be uncorrelated with the unobserved severity of illness,⁷ but correlated with the endogenous choice variables. Thus, the distance from each patient in our sample to Cedars-Sinai Medical Center (for instance) is an appropriate instrument. We use this insight to construct instruments based on the distance from each patient to each hospital. We then estimate the patient mortality hazard as a function of the choice of hospital and patient-level data, using these distance measures to instrument for the endogenous choice of hospital.

For our study, we use discharge data from the State of California Office of Statewide Health Planning and Development (OSHPD) and hospital characteristic data from OSHPD and the American Hospital Association (AHA) over the period 1989-1994.⁸ We focus on Southern California Medicare enrollees with a diagnosis of pneumonia. We examine only Medicare enrollees as the standard Medicare insurance program allows enrollees access to almost all hospitals and the enrollees incur the same out-of-pocket expenditure regardless of the hospital to which they are admitted. Thus, there will be less unobserved heterogeneity in their choice of

⁵ To analyze whether this assumption is sensible, it is perhaps useful to consider an instance where it would be invalid. For instance, suppose that nursing homes had more severely ill patients than their surrounding population and tended to locate near high quality hospitals. Then, if we did not observe whether patients were admitted from nursing homes, unobserved severity of illness would be higher near high quality hospitals and our instruments would be invalid. Fortunately, we can observe and eliminate patients who are admitted from nursing homes.

⁶ Luft, et al. (1990) show that distance is an important predictor of hospital choice.

⁷ It is important to note that while distance to a given hospital is a good instrument, distance to the *chosen* hospital will certainly be correlated with unobserved severity and hence be endogenous.

⁸ We thank OSHPD for making this data available to us.

hospital, which should lead to our instruments being more highly correlated with the endogenous hospital choice variable and hence to more precise results.

We examine pneumonia diagnoses for several reasons. First, it is a common affliction. Pneumonia is the fourth leading cause of death among the elderly, and it is the most common cause of death due to an infectious disease.⁹ Second, changes in hospital procedures can affect pneumonia outcomes; thus pneumonia mortality rates contain a component that reflects the quality of care. Additionally, hospital-specific pneumonia mortality rates have often been used as a benchmark of hospital quality. Our results should provide some indication as to whether the estimates of hospital quality formulated in these papers suffer from selection bias.

We find that the IV estimation yields significantly different coefficients on mortality from standard GLS estimation, even after controlling for demographic information and comorbidities. The importance of using IV is verified by Hausman specification tests which reject the consistency of the GLS estimates in five out of the six years and in the pooled years estimates. We interpret these results as evidence that unobserved patient severity is correlated with the choice of hospital, thereby inducing bias in the estimates of adjusted hospital mortality. In order to determine whether our model and instruments are appropriate, we perform a test of the overidentifying restrictions of the model, as suggested by Hansen (1982). We fail to reject the null hypothesis of an appropriate model for all years. That is, the data supports our assumption that the distance instruments are exogenous. In addition, we construct tests for weak instruments as suggested by Staiger and Stock (1997). Our instruments appear to be good ones in the sense that they are highly correlated with the choice of hospital and are not weak instruments. We note one important caveat to our results. The standard errors are too high to be able to accurately rank individual hospitals.

We are also interested in whether failure to correct for the selection bias would lead to different, and presumably incorrect, inferences regarding the correlates of hospital quality. We therefore examine the relationship between the estimated hospital mortalities and the characteristics of hospitals for both the IV and GLS estimates. We find that the estimated

relationships between hospital characteristics and mortality rates differ between the IV and GLS estimators. The differences between the IV and GLS estimates of hospital quality are material. Depending on the method used to estimate hospital quality, one would make different inferences regarding the correlates of hospital quality. Unless researchers control for unobserved patient severity they risk making incorrect inferences regarding the determinants of hospital quality.

The remainder of the paper has the following structure. Section 2 describes the model. Section 3 discusses the data we use in the analysis. Section 4 presents the results. Section 5 provides some implications of the results. Section 6 concludes.

2. The Model

We adopt a discrete-time duration model specification of in-hospital mortality. There are two types of equations to our model: a mortality hazard equation and a set of hospital choice equations, one for each hospital. We use the estimated coefficients from the mortality hazard to analyze the factors that are correlated with quality. We detail our estimation below.

Model of mortality and hospital choice

Our model of hospital choice is as follows: once an individual becomes ill with pneumonia, the patient and her physician determine if she should be hospitalized. After this determination is made, the patient (and/or the physician acting as the patient's agent) chooses the hospital at which the patient will receive care. We assume that the choice of hospital is based on the perceived quality of the treatment that the patient would receive at the hospital, the patient's

⁹ National Center for Health Statistics (1996).

severity of illness, the cost to the patient of the treatment and the distance the patient must travel to the hospital.¹⁰

For patient i and hospital j , the hospital choice model can be expressed as:

$$(1) \quad c_{ij} = f_j(z, \delta_i, u_{ij}),$$

where c_{ij} is a dummy variable that indicates that patient i has chosen hospital j ; z is the set of observable characteristics for all hospitals that might include locations, service offerings and ownership type; δ_i is the set of observed patient characteristics that includes location; and u_{ij} is the unobserved (to the researcher but not the patient/physician) component of choice. The choice model (1) is indexed by j , where j takes on values from $1, \dots, J$, the number of hospitals in Southern California. Thus, there are J choice equations. While the hospital choice equation is an integral part of our model, we do not estimate it. Instead, we use single equation instrumental variables techniques to estimate the hospital mortality equation without estimating the choice equation.

We model in-hospital mortality by estimating the hazard for death each day, which is the probability of a patient dying in the hospital on a given day conditional on being alive at the start of that day. We assume that the hazard depends on the observed and unobserved severity of illness of the patient, the number of days since admission, and the quality of the treatment received at the hospital chosen. We postulate a linear relationship between the hazard and these predictors. The mortality equation is:

$$(2) \quad m_{it} = \beta c_i + \gamma x_i + \alpha d_t + s_{it} + \varepsilon_{it},$$

where m_{it} is a dummy variable that denotes mortality on the given day after admission; c_i is the vector of dummy variables (c_{i1}, \dots, c_{iJ}) that indicate the choice of hospital; x_i is a vector of patient characteristics that can affect mortality, including age, race, sex, and disease stage; d_t is a vector of dummy variables that denotes the number of days since admission; s_{it} is the

¹⁰ Luft *et al.* (1990) shows that distance and quality are significant predictors of hospital choice.

unobserved (to the researcher but not necessarily to the patient/physician) severity of illness; and ε_{it} is the residual component of mortality. The parameter vectors to estimate are β , γ , and α . The vector β specifies a separate quality for each hospital, and it is of primary interest.

Recall that (2) specifies the probability of dying in the hospital on a given day, conditional on being alive. To account for the conditional nature of the hazard, we keep observations for each patient as long as the patient is alive and in the hospital. As we do not observe post-discharge death, a live discharge is treated as a censored observation. For instance, if a patient dies on the 15th day after the date of admission, we would have 16 observations for that patient. Similarly, if a patient were discharged alive on the 9th day after the date of admission, we would have 10 observations for that patient. Because we specify a dummy variable for each day following the date of admission, we truncate observations after 30 days, to limit the number of regressors. Thus, a patient with a length of stay of more than 30 days is also treated as a censored observation, and the estimation process only uses the information that the patient survives at least 30 days in the hospital.

We assume that unobserved severity of illness s_{it} and the purely random error ε_{it} in (2) are identically and independently distributed in the population and that they are orthogonal to the regressors x_i and d_i . Recall that we have controlled for income, race, age, sex, diagnosis and comorbidities, and thus it is the *residual* health status that is assumed to be distributed identically in the population.

However, patients may choose hospitals differentially based on their unobserved severity of illness. For instance, patients with a high unobserved severity of illness may be more likely to choose a hospital that is of high quality. Econometrically, this will result in a correlation between the error terms u_{ij} in the choice equations (1) and the error term $s_{it} + \varepsilon_{it}$ in the mortality equation (2). For instance, if hospital j is of high quality and hospital k is of low quality, then s_{it} may be positively correlated with u_{ij} but negatively correlated with u_{ik} . Thus, the hospital choice dummy variables, c_i will be correlated with the error term s_{it} and hence be endogenous, and the estimates of β and γ from a standard model will not be consistent.

We use geographic variables as instruments for the endogenous hospital choices, and then estimate the mortality hazard with instrumental variables. An example of one of our instruments is the distance from each patient to hospital j . As we assume in our choice model (1), there is considerable evidence that distance to a hospital is negatively correlated with the choice of hospital (e.g. Luft *et al.* (1990) and Burns and Wholey (1992)). Thus, the distance from the patient to hospital j will be highly correlated with the dummy variable c_{ij} . While this distance is correlated with the hospital choice it will be uncorrelated with the error term, by our assumption that the unobserved component of severity of illness is identically distributed in the population.

It should be noted that according to our model, patients who are more severely ill may choose to travel further to seek higher quality care. Thus, distance to the *chosen* hospital will be correlated with the severity of illness of the patient, and hence will not be a valid instrument.

Choice of instruments

As any function of distance would be an appropriate instrument, there are infinitely many possible variables related to distance that we could use as instruments. Since we have $J - 1$ endogenous variables (one for each hospital minus a constant) we require at least this many instruments to identify the model. In order to capture the non-linear effect of distance, we decided to use $2J$ instruments: J instruments that indicate the distance d_{ij} in kilometers between patient i and hospital j and J more instruments of the form $\exp(-\phi d_{ij})$, where ϕ is a parameter. Because patients will tend to choose hospitals that are close to their homes, we set these instruments to 0 for any hospital that is located more than 120 km from the patient.

While the latter set of instruments would be consistent given any value of ϕ , we estimate ϕ in order to increase the efficiency of our results. The ideal method to estimate ϕ is in a joint estimation of all of the parameters using a multi-equation system of the choice and mortality equations. However, this is not feasible, since ϕ has a non-linear functional form given our instruments, and non-linear least squares for this large a problem is computationally too difficult. Thus, we estimate ϕ with a series of single equation non-linear regressions for the choice model (1). We use a very simple functional form of (1):

$$(3) c_{ij} = \delta_{1j} + \delta_{2j}d_{ij} + \delta_{3j} \exp(-\phi d_{ij}) + u_{ij},$$

and estimate (3) separately for each hospital in the data. In our estimation of (3), we obtained mean coefficients (across hospitals) for ϕ of 0.2 to 0.3 for different years in the data. Thus, we set $\phi=0.25$ for all of our estimation results.

Efficient estimation and testing

We note that the hazard of the mortality equation in (2) is a linear probability model. With the linear probability model, the β coefficients have a straightforward interpretation: they are the incremental probabilities of death on any day from seeking treatment at a particular hospital. The reason that we use a linear probability model instead of a more common Weibull or Lognormal specifications for the hazard is that it is extremely difficult to use non-linear models such as these with endogenous variables. To estimate a Weibull or Lognormal model, one must evaluate the probability of death on a given day, conditional on parameter values. For our case, this probability (of death for each patient) can only be computed by integrating over the joint density of the endogenous variables and the mortality equation error term. As the endogenous choice variables are derived from the patient choice equation (1), we would have to integrate over all the error terms in (1), as well as the mortality equation error term. Computation of such models is not feasible due to the dimensionality of this integration.

Because errors in a linear probability model are heteroskedastic, we use a GLS procedure. While we could use a general heteroskedasticity correction as suggested by White (1980) and White (1982), we use the easily derived asymptotically optimal weighting matrix which accounts for the fact that we know the exact functional form of the heteroskedasticity. This allows us to derive an asymptotically efficient estimator, which is necessary to perform the Hausman test of endogeneity described below. The heteroskedasticity correction is performed in the same way for both the standard GLS estimates and the IV estimates.

The idea of our correction is as follows. If we knew the true value of the error term $\eta_{it} \equiv s_{it} + \varepsilon_{it}$, then the fact that m_{it} is a binomial ensures that:

$$(4) \quad \text{Var}(\eta_{it}) = \eta_{it}(1 - \eta_{it}).$$

We could then correct for heteroskedasticity by weighting each observation by the inverse of its standard deviation.

Since we do not know the true values of η_{it} , we use a two-stage process. We first perform an initial regression (of OLS or non-heteroskedasticity corrected IV, depending on the case). As this regression is consistent, it provides consistent estimates $\hat{\eta}_{it}$ of the residuals η_{it} . We then weight each observation according to its estimated standard deviation, so that the weighting matrix is a diagonal matrix with diagonal elements of

$$(5) \quad W_{it,it} = 1/\sqrt{\max\{\hat{\eta}_{it}(1 - \hat{\eta}_{it}), 0.01\}}.^{11}$$

By pre-multiplying all of the variables in (2) by W , the problem is asymptotically transformed into a standard homoskedastic problem. To complete our estimation, we perform a second stage regression, where we transform the variables in (2) (mortality, hospital choice and demographics) using (5) and then estimate OLS or IV (depending on the case) on the transformed data. This will provide us with asymptotically efficient estimates that exploit the exact functional form of our heteroskedasticity. Note that it is neither necessary nor optimal to transform the instruments. We then compute an estimated variance for these second stage parameter estimates using the standard IV or OLS formulae on the transformed data.

To evaluate our estimates, we perform a Hausman (1978) specification test for the endogeneity of hospital choice. The Hausman test statistic is given by:

$$(8) \quad H = (\hat{\theta}^{IV} - \hat{\theta}^{GLS})' (\text{EstVar}\hat{\theta}^{IV} - \text{EstVar}\hat{\theta}^{GLS})^{-1} (\hat{\theta}^{IV} - \hat{\theta}^{GLS}),$$

where $\hat{\theta} = (\hat{\beta}, \hat{\gamma}, \hat{\alpha})$. Under the null hypothesis of no endogeneity, H is distributed χ^2 with degrees of freedom given by the number of included endogenous variables ($J-1$). We also

¹¹ The “max” is necessary for cases where the predicted residual is negative or greater than one in order to avoid a negative square root. Asymptotically, these cases will tend to disappear and hence not affect the results.

perform a test of the overidentifying restrictions, as suggested by Newey (1985), Hansen (1982) and Bowden and Turkington (1984). This test statistic is given by:

$$(9) \quad O = s^2(Z'e)'(Z'Z)^{-1}(Z'e),$$

where Z is the matrix of instruments, e is the vector of error terms from the second-stage IV estimation routine and s^2 is the estimated variance of the error terms from this regression. Under the null hypothesis of exogeneity of the instruments, O is distributed χ^2 with degrees of freedom given by the number of overidentifying restrictions ($J-I$).

3. The Data

The primary source of data that we use is the Version B Discharge Data from the State of California, Office of Statewide Health Planning and Development (OSHPD). The data lists records for all patients discharged from any California acute-care hospitals during 1989-1994. We used the data to estimate the mortality model separately for each of the six years in the data, and jointly, pooled together for all six years.

For our purposes, the Version B data provides patient level information on zip code of residence, up to five ICD-9-CM disease codes, race, sex, age (by classes), hospital that the patient was admitted to, method of admittance (emergency room, etc.) and disposition (normal discharge, death, etc.). From this data we kept patients age 65 or older with an ICD-9-CM code indicating a diagnosis of pneumonia¹² and who were admitted to a hospital in Los Angeles, Orange, San Bernardino, or San Diego counties. We removed from the data set any patient whose source of admission was other than the emergency room or routine. Thus, we removed all patients who were transferred into the hospital from another medical facility and those who were admitted from an intermediate care or skilled nursing facility. The reason for this is that intermediate care facilities and nursing homes may locate near higher quality hospitals, and thus

¹² We use ICD-9-CM codes for pneumonia of 48.1, 48.2, 48.5, 48.6 and 48.38, as suggested by Iezzoni (1994). In order to not eliminate pneumonia patients whose primary diagnosis was incorrectly coded as something other than pneumonia, we keep all patients who have one of these codes among the top five disease codes specified.

the assumption that residual severity of illness is identically distributed in the population would be violated for these patients. We could attempt to control for admittance from a skilled nursing facility. However, as some of these facilities are affiliated with hospitals, patients who are admitted from skilled nursing facilities will have very different choice sets.

As we estimate separate intercepts for each hospital, including small hospitals in our sample would decrease the efficiency of our overall estimates. Accordingly, we eliminated hospitals that treated less than 80 Medicare pneumonia patients for a given estimation period. We then merged the California discharge data with the OSHPD Annual Report of Hospitals and the American Hospital Association (AHA) Annual Survey of Hospitals databases. For our purposes, the AHA survey provides hospital-level data on the address and characteristics (e.g. number of beds and physicians, whether they are a teaching hospital, etc.) of each hospital.

We restrict our attention to pneumonia patients for several reasons. First, there is evidence that severity adjustment mechanisms work best when they are disease specific.¹³ Therefore, we focus on one disease. Pneumonia was chosen because it is extremely prevalent in the Medicare population: we have over 28,000 patients for each year of the analysis. In-hospital death is a relatively frequent outcome for pneumonia patients (the in-hospital mortality rate in 1989 was 17.9%), which makes it an appropriate disease to examine given that we have only the mortality status at the time of discharge. Pneumonia is the leading cause of death among infectious diseases, the sixth leading cause of death in the US, and the fourth leading cause of death for those over 65 years of age (National Center for Health Statistics (1996)).¹⁴ Importantly, there is evidence that the appropriately adjusted in-hospital mortality rate for pneumonia is correlated with the quality of care provided at that hospital.¹⁵ Thus, if we can consistently estimate hospital mortality for pneumonia, then we are obtaining information about the quality of care being provided to pneumonia patients and perhaps even to other patients at a given hospital.

¹³ See Wray *et al.* (1997).

¹⁴ Pneumonia is also the leading cause of death among patients with nosocomial (hospital acquired) infections (Pennington (1994)).

¹⁵ See Keeler *et al.* (1990) and McGarvey and Harper (1993).

Moreover, unlike stroke and heart attack diagnoses, pneumonia does not place a high premium on the time until care is rendered. This suggests that individuals in our data set have a substantial amount of discretion over their choice of hospital. Thus, our selection of pneumonia allows us to assess the potential importance of selection bias in measuring hospital quality using those diagnoses in which patients have a significant amount of choice over the hospital where they will receive care. Additionally, we can exclude distance to the chosen hospital as a predictor of mortality, since the incidence of adverse medical outcomes from extra travel distance will be negligible.

Lastly, several research efforts have attempted to use inpatient mortality from pneumonia as the sole diagnosis or part of a group of diagnoses to measure hospital quality (e.g. Whittle, *et al.* (1998), Rosenthal, *et al.* (1998), Iezzoni, *et al.* (1996), Hartz *et al.* (1989), Shortell and Hughes (1988) and Dubois, *et al.* (1987)). Thus, our findings will provide some indication on the extent of the biases that may plague analyses that do not correct for selection.

We note that the admission to any hospital for pneumonia may have a discretionary element that could lead to biases induced by the treatment decision. However, as long as location is orthogonal to admission decisions, our estimates should not suffer from biases due to differential admissions. The evidence that exists on this point indicates that the hospitalization decision among Medicare patients with pneumonia is primarily determined by the severity of illness as opposed to demographic factors which will be correlated with location (Fine *et al.* (1997)). This suggests that our location instruments should remain exogenous even with differential admissions.

We want our model of mortality to be the best predictor of death possible in a linear framework. We follow the suggestions of Wray *et al.* (1997) and Iezzoni (1994) in constructing the set of explanatory variables, x_i , given the constraints of our data. Wray *et al.* (1997) advise that appropriate adjustment mechanisms must control for the principle diagnosis within a DRG, contain demographics as proxies for pre-existing physiological reserve, and measure the number and severity of comorbidities. Following Iezzoni (1994), we control for the principle diagnosis

and the measure of comorbidities using dummies for the Disease Stage.¹⁶ Disease Staging maps from the list of comorbid ICD-9 codes to a severity scale that ranges from 1-4 where stage one is the least severely ill and stage four is death. Disease Staging has been shown to be as good as some risk adjustment mechanisms that require data collected from medical records.¹⁷ As some stages within the Disease Staging contain very few observations, we group the Disease Stages into five categories; the exact categories are given in Table 1. Besides Disease Stage, the other observed demographics that we use to control for severity are race dummies, age dummies, sex dummy, emergency admittance dummy, estimated income, and estimated income squared.

As noted above, all of these variables, except income, are readily available from the OSHPD data. Excluding income could be problematic, as income is a potentially important predictor of severity of illness since it may affect the prior health care, nutrition and fitness of patients and it is not orthogonal to location. Fortunately, we can use 1990 Census data to proxy for patient income. Recall that the discharge data indicates the zip code, race and age class of the patient.¹⁸ From the census, we obtain mean household income for individuals with the given zip code, race and age class, and substitute this for the patient income. For patients who were from zip codes that have been created after the 1990 Census, we used income information from the closest zip code that existed at the time of the 1990 Census.

In order to construct our instruments, we need to have data on the distance from each patient's home to each hospital. We approximated both the locations of the hospital and of a patient's residence using zip code information. There are some exceptions to this approximation methodology and we detail those below. We obtained patient zip codes from the discharge data and hospital zip codes from the AHA data. We then used the Census TIGER database to find the latitude and longitude of the centroid of each zip code. Given the latitudes and longitudes, we computed the distance from any two locations using the standard great circle trigonometric formulas.

¹⁶ See Gonnella, Hornbrook, and Louis (1984).

¹⁷ See Thomas and Ashcroft (1991).

¹⁸ Geronimus *et al.* (1996) empirically analyze the validity of census information as a proxy for income.

There was one exception to this location method. Many zip codes have more than one hospital located inside them. If we computed location solely by the centroid of the zip code, the distance from every patient to each of these hospitals would be the same. We would then have the instruments that correspond to each of these hospitals be exactly the same, and hence be perfectly collinear. We solved this problem by obtaining more detailed address-level geographic data for hospitals located in zip codes with more than one hospital. We also obtained the address-level geographic data from the Census TIGER database, which stores the geographic location of every block corner and will interpolate from that to find the location of any address.

As we constructed our data set by merging information from several sources, we had to eliminate a certain number of the observations due to matching errors. First, we eliminated patients whose race was listed as ‘other’ or ‘unknown’ because of the impossibility in obtaining income data given our method. Second, we eliminated patients with miscoded zip codes or zip codes from outside of California. Third, as mentioned earlier, we omitted patients whose source of admission is other than routine or emergency room. Finally, we eliminated patients from hospitals that treated under 80 Medicare pneumonia patients. We were still able to keep the majority of patients (72%): in total for all six years we started with 248,816 pneumonia patients with age over 65 discharged from Southern California hospitals, and our final data sets for separate year estimation had 178,972 observations.

Table 1 lists the population characteristics of our sample. In 1989, the typical patient is a white female between the ages of 75 and 80 who was admitted via the emergency room with a Disease Stage of 1.1. The estimated average income is \$27,629. By 1994, our data set included a greater percentage of minorities, and (1990 Census) income had risen slightly to \$28,135. Note also that in-hospital 30-day mortality is quite close to in-hospital mortality (16.2% vs. 17.9% for 1989; 12.5% vs. 13.0% for 1994), thus we do not lose much information by censoring at 30 days.

At the end of Table 1, we provide some of the distance characteristics for our sample. Patients generally choose hospitals that are very nearby, although this effect has been diminishing over time: the average distance traveled to receive care increased from 9.5 kilometers to 11.8 kilometers between 1989 and 1994. We also include the mean coefficients

from the choice on distance regression of (3). As reported earlier, the mean coefficients on ϕ are roughly 0.25. Also, the coefficients on δ_{2j} and δ_{3j} show that distance is an important predictor of choice. The fact that δ_{2j} changes in sign from 1989 to 1994 but that δ_{3j} increases shows that the effects of distance are becoming more non-linear over time. Thus, by 1994, patients have a relatively higher elasticity of distance for nearby hospitals, but are still very unlikely to travel large distances.

4. Results

We estimate the parameters from equation (2) using the IV methods separately for each year in our data and pooling all of the years into one data set. For comparison purposes, we also estimate the same regressions without controlling for the endogeneity of hospital choice. We denote these regression results GLS. Table 2 displays the results of several specification tests for the GLS and IV models. Our primary finding is that the specification tests indicate that GLS estimates of hospital quality are biased and the IV estimates apparently correct for this bias. The Hausman test rejects, at the 1% level, the hypothesis that hospital choice is exogenous in every year and for the pooled year regression. We also tested the overidentifying restrictions of the IV estimator as suggested by Hansen (1982)—we designate this the Hansen test. The Hansen test fails to reject the hypothesis that our model is correctly specified for each year at the 5% level. We take these two test results as strong evidence in favor of our proposed methodology.

In order for IV methodology to generate consistent estimates the instruments must be correlated with the endogenous variables. Additionally, Bound, Jaeger and Baker (1995) and Staiger and Stock (1997) show that if the instruments are only weakly correlated with the explanatory variables, the coefficients may be biased, in large, finite samples. We tested for weak instruments by calculating a Wald test of the null hypothesis that the true relationship between

the instruments and the choice of hospital is zero, as suggested by Staiger and Stock (1997). We resoundingly reject the null for all six years and the pooled regression.

Figure 1 is the scatterplot of the GLS and IV point estimates of hospital quality from the separate year regressions normalized so the mean is zero. The spread of the IV estimates is much greater than the spread of the GLS estimates indicating that the GLS methodology underestimates the actual distribution of hospital quality. This, in turn, suggests that good hospitals do attract sicker patients. There is a small positive correlation ($\rho = 0.14$) between the two sets of estimates, however GLS estimates explain very little of the variance in the IV estimates. The R^2 of the regression of the GLS on the IV point estimates of hospital quality is 0.02. In other words, the GLS and IV point estimates of hospital mortality are quite different. For example, of the 51 hospitals that are significantly different (at the 10% level) from the average hospital according to the GLS estimates in the pooled years regression, only 19 (37%) of these hospitals are also significantly different than the average hospital according to the IV estimates of quality. This indicates that those previous studies that used pneumonia to measure hospital quality likely suffer from significant bias and misclassification.

Table 3 presents the coefficient estimates of the demographic variables for the IV and GLS pooled years regressions. The GLS estimates are generally more precise than the IV estimates. For the most part, we cannot reject the hypotheses that the IV demographics coefficient estimates are different from the GLS estimates. The IV results indicate that being younger, female, Asian, a non-emergency room admit and less severely ill according to the Disease Staging methodology decrease the likelihood of dying. The income coefficients were insignificant at traditional confidence levels.

The IV quality estimates are somewhat imprecise. The average of the absolute value of the t-statistic for the hospital quality coefficients in the IV estimation is 1.04 for the yearly regressions and 1.25 for the pooled years regression. Thus, our methodology currently is able to

identify only a minority of the hospitals as significantly different than the average hospital at traditional confidence levels. In the yearly regressions, 20% of the hospitals are significantly (at the 10% level) different than the average hospital, while 28% of the hospitals on the pooled years regression are significantly better or worse than the average hospital.

5. Implications of the Estimates of Hospital Quality

The findings presented above indicate that failing to correct for unobservable severity is likely to lead to substantial bias in measuring hospital mortality. In this section we explore whether the IV results differ from the GLS results in ways that might affect interpretations regarding the correlates of hospital quality. While the precision of the individual hospital coefficients is such that we would hesitate to rely on the individual hospital point estimates, we feel confident in examining general population properties based on the estimated parameter values. Accordingly, we regress our estimated IV qualities from the separate year regressions against factors that might be correlated with hospital quality. For comparison purposes, we also regress our estimated GLS qualities against these factors.

As the estimates from the yearly regressions form a panel we use a random-effects model to estimate the coefficients.¹⁹ The list of explanatory variables available to us includes dummy variables for type of organization that operates the hospital (public, not-for-profit, and for-profit), dummy variables indicating whether the hospital is a member of the Council of Teaching Hospitals, whether the hospital operates a long-term care facility, and whether the hospital operates a geriatric care unit. We also included the size of the hospital as measured by staffed

¹⁹ Note that in this section, our dependent variables is a coefficient estimates of the hospital intercepts in (2). Since we know the distribution of those estimates via the variance-covariance matrix we could potentially increase the efficiency of our random-effects model by incorporating this information in the estimation. We experimented with this, but were unable to accurately estimate the variances of the random effects model in this case. Instead, we present standard errors that are robust to heteroskedasticity, for all

beds interacted with the for-profit status of the hospital to the list of explanatory variables. The idea is that the type of organization may affect the relationship between hospital size and quality. Finally, we include the hospital's average length of stay across Medicare, pneumonia patients and its occupancy rate to the list of left-hand side variables. We acquired the hospital characteristics from the AHA data set, except for ownership, bed size, length of stay and utilization rate, which are from the OSHPD Annual Report of Hospitals data set.

We estimate the parameters for two different sets of regressions. The independent variables for the first set of regressions only includes the dummy variables for the type of organization that operates the hospitals. For each set of regressions we estimate the parameters using both the IV and GLS estimates of hospital quality. The second set of regressions includes all of the explanatory variables given above in the list of independent variables.

Table 4 presents the estimation results of the random-effects model. The results indicate that the differences between the IV and GLS estimates of hospital quality are material. One would make different inferences regarding the correlates of hospital quality using different methods to estimate hospital quality. That is, unless researchers control for unobserved patient severity, they risk making incorrect inferences regarding the determinants of hospital quality.

For instance, according to the IV estimates listed in column (2) admission to a typical private, not-for-profit hospital will significantly (at the 10% level) reduce the likelihood of death relative to for-profit hospitals. The expected daily mortality is 0.6 percentage points lower in private, not-for-profit hospitals than in for-profit hospitals and 0.3 lower than public hospitals although this later difference is not significant. However, according to the GLS estimates (column (4)) there is no significant difference between the typical public, private, not-for-profit or for-profit hospital. These results suggest that patients appear to factor their unobserved

the results of this section. While the standard errors are not robust to autocorrelation, we correct for this via a random effects model.

severity of illness and the hospital's quality into their admission decision, and this selection process implies that the typical private, not-for-profit hospital treats patients with higher unobserved severity than a typical public or for-profit hospital.

Also, the IV regression results reported in column (3) of Table 4 control for other hospital characteristics. The estimates imply that the quality of for-profit hospitals increases with number of beds, while there are no benefits to size for the other types of hospitals. An increase of 10 beds is expected to reduce the daily mortality rate in for-profit hospitals by 0.7 percentage points. The corresponding GLS estimates (column (5)) indicate that only public hospitals increase their quality by increasing their size. The correlation between occupancy rates and mortality rates differs between the IV and GLS estimates. According to the IV estimates, an increase in the occupancy rate is correlated with a reduction in the expected mortality rate. The coefficient in the GLS regression is not significant.

There are also some similarities in the coefficient estimates and their significance across the estimation methodologies. Hospitals with higher average length of stays have lower expected mortality rates in both sets of estimates. Increasing the average length of stay by one day is expected to reduce the daily mortality rate by approximately 0.2 percentage points according to the IV estimates. Also, the coefficients on teaching hospitals and hospitals with geriatric care and long term care units were insignificant in both regressions.

6. Conclusions

In this study, we have examined the quality of care for hospitals, controlling for unobserved patient severity of illness. We postulate a model of patient choice and a hazard model of mortality. Then, we show that we can consistently estimate our mortality model within an IV framework given geographic data and the assumption that unobserved (after controlling for demographics) severity of illness is identically distributed in the population. We estimate our

model for Medicare pneumonia patients in California hospitals from 1989 to 1994. We find that controlling for unobserved severity with an IV estimator results in very different predictions for hospital quality than standard regression techniques. Econometric tests suggest that the GLS estimates of hospital quality suffer from selection bias and that the IV model is appropriately specified. The results indicate that the differences between the IV and GLS estimates of hospital quality are material. One would make different inferences regarding the correlates of hospital quality based upon the method used to estimate hospital quality. The methods outlined here potentially have wide applicability in the analysis of the causes and consequences of hospital quality. Inquiries into understanding the determinants of hospital quality obviously will need to develop unbiased measures of hospital quality. The methods we suggest here provide a way to construct such measures for many widely available data sets.

References

- Bound, J., D. A. Jaeger, and R. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association* 90, 443-450.
- Bowden, R. and D. Turkington, 1984, *Instrumental variables* (Cambridge University Press, Cambridge).
- Burns, L. and D. Wholey, 1992, The impact of physician characteristics in conditional choice models for hospital care, *Journal of Health Economics* 11, 43-62.
- Dubois, R., Brook, R., and W. Rogers, 1987, Adjusted hospital death rates: A potential screen for quality of medical care, *American Journal of Public Health* 77, 1162-6.
- Fine, M.J., et al., 1997, The hospital admission decision for patients with community-acquired pneumonia: Results from the pneumonia patient outcomes research team cohort study, *Archives of Internal Medicine* 157, 36-44.
- Geronimus, A., J. Bound, and L. Neidert, 1996, On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics, *Journal of the American Statistical Association* 91, 529-37.
- Gonnella, J.S., W.C. Hunt, and D.Z. Louis, 1984, Staging of disease: A case-mix measurement, *Journal of the American Medical Association* 251, 637-44.
- Hansen, L, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-54.
- Hartz, A. et al., 1989, Hospital characteristics and mortality rates, *The New England Journal of Medicine* 321, 1720-1725.
- Hausman, J., 1978, Specification tests in econometrics, *Econometrica* 46, 1251-1271.
- Iezzoni, L.I., ed., 1994, *Risk adjustment for measuring health care outcomes* (Health Administration Press, Ann Arbor, MI).
- Iezzoni, L., M. Shwartz, A. Ash, J. Hughes, J. Daley, and Y. Mackiernan, 1996, Severity measurement methods and judging hospital death rate for pneumonia, *Medical Care* 34, 11-28.
- Keeler, E, L. Rubenstein, K. Kahn, D. Draper, E. Harrison, M. McGinty, W. Rogers, and R. Brook, 1992, Hospital quality and quality of care, *Journal of the American Medical Association* 268, 1709-1714.
- Keeler, E., K. Kahn, D. Draper, M. Sherwood, L. Rubenstein, E. Reinisch, J. Kosecoff, and R. Brook, 1990, Changes in sickness at admission following the introduction of the prospective payment system, *Journal of the American Medical Association* 264, 1962-1968.

- Luft, H., et al., 1990, Does quality influence choice of hospital? *Journal of the American Medical Association* 263, 2899-2906.
- McClellan, M., B. McNeil, and J. Newhouse, 1994, Does more intensive treatment of acute myocardial infraction in the elderly reduce mortality-analysis using instrumental variables, *Journal of the American Medical Association* 272, 859-866.
- McClellan, M. and J. Newhouse, 1997, The marginal cost-effectiveness of medical technology: A panel instrumental-variables approach, *Journal of Econometrics* 77, 39-64.
- McGarvey, R. and J. Harper, 1993, Pneumonia mortality reduction and quality improvement in a community hospital, *Quality Review Bulletin* 19, 124-30.
- National Center for Health Statistics, 1996, Available on the Internet at <http://www.cdc.gov/nchswww/nchshome.htm>.
- Newey, W., 1985, Generalized method of moments specification testing, *Journal of Econometrics* 29, 229-256.
- Pennington, J., 1994, *Respiratory infections: Diagnosis and management*, 3rd ed. (Raven Press, New York).
- Rosenthal, G., A. Shah, L. Way, and D. Harper, 1998, Variations in standardized hospital mortality rates for six common medical diagnoses, *Medical Care* 36, 955-964.
- Shortell, S., and E. Hughes, 1988, The effects of regulation, competition, and ownership on mortality rates among hospital inpatients, *The New England Journal of Medicine* 318, 1100-1107.
- Staiger, D. and J. H. Stock, 1997, Instrumental variables regression, *Econometrica* 65, 557-586.
- Thomas, J.W. and M.L.F. Ashcroft, 1991, Measuring severity of illness: Six severity systems and their ability to explain cost-variations, *Inquiry* 28, 39-55.
- United States General Accounting Office, 1994, Health care reform: 'Report cards' are useful but significant issues need to be addressed (The White House Domestic Policy Council, Washington, DC).
- White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817-38.
- White, H., 1982, Instrumental variables regression with independent observations, *Econometrica* 50, 483-500.
- Whittle, J. et al., 1998, Relationship of provider characteristics to outcomes, process, and cost of care for community-acquired pneumonia, *Medical Care* 36, 977-987.

Wray, N., J. Hollingsworth, N. Petersen, and C. Ashton, 1997, Case-mix adjustment using administrative databases: A paradigm to guide future research, *Medical Care Research and Review* 54, 326-356.

Table 1

Inpatient Population Means for Various Characteristics: 1989 and 1994

Characteristic	1989	1994
Patients	29,278	30,838
In-Hospital Mortality	17.9%	13.0%
In-Hospital 30-day Mortality	16.2%	12.5%
Mean Estimated Income	\$27,629	\$28,135
Age 65-69 Years	14.9%	14.6%
Age 70-74 Years	18.5%	18.5%
Age 75-80 Years	20.7%	20.9%
Age 80-84 Years	20.0%	20.0%
Age 85 Years and Older	25.9%	26.0%
Female	53.8%	53.7%
White	79.9%	72.7%
African American	6.2%	6.7%
Hispanic	10.0%	13.2%
Asian	3.8%	7.2%
Native American	0.1%	0.2%
Emergency Room Admittance	70.3%	68.1%
Disease Stage 1.1 (death rate)	75.3% (0.08)	74.9% (0.06)
Disease Stage 1.3-2.3 (death rate)	6.8% (0.08)	6.1% (0.08)
Disease Stage 3.1-3.6 (death rate)	6.5% (0.19)	6.9% (0.17)
Disease Stage 3.7 (death rate)	9.7% (0.21)	10.8% (0.21)
Disease Stage 3.8 (death rate)	1.7% (0.50)	1.3% (0.51)
Average Distance Traveled to Hospital	9.45 km	11.8 km
Mean value of δ_{2j}	-0.00011	0.00176
Mean value of δ_{3j}	0.587	2.757
Mean value of ϕ	0.260	0.258

Table 2
Specification Test Statistics for Separate and Pooled Year Regressions.

	1989	1990	1991	1992	1993	1994	All Years
Hausman Statistic (p-value)	244.1 (0.00)	404.3 (0.00)	223.8 (0.00)	248.8 (0.00)	230.6 (0.00)	209.3 (0.00)	754.9 (0.00)
Hansen Statistic (p-value)	148.3 (0.48)	141.6 (0.52)	110.9 (0.96)	125.8 (0.82)	117.4 (0.90)	159.0 (0.11)	222.1 (0.10)
Wald Test of Weak Instruments (p-value)	189,926 (0.00)	165,153 (0.00)	154,892 (0.00)	146,932 (0.00)	134,692 (0.00)	128,438 (0.00)	745,010 (0.00)
Number of Hospitals	147	142	139	142	137	137	194
Patients	29,278	28,626	29,233	30,580	30,417	30,838	192,386

Note: The Hausman, Wald, and Hansen statistic are all distributed χ^2 . The number of hospitals minus one gives the degrees of freedom for the Hausman and Wald statistics. The degrees of freedom for the Hansen test is given by the number of hospitals plus one.

Table 3
Estimates of the Effect of Demographic Variables on Mortality
from GLS and IV methods: Pooled Years Regression
(Standard errors in parentheses)

Variable	IV Coefficients	GLS Coefficients
Income/100	0.0053 (0.0033)	0.0032 (0.0022)
(Income/1000) ²	-0.00025 (0.00032)	-0.00022 (0.00024)
Age 70-74 Years (Age 65-69 omitted)	0.00018 (0.00037)	0.00051 (0.00025)
Age 75-80 Years	0.0017 (0.00043)	0.0017 (0.00026)
Age 80-84 Years	0.0041 (0.00051)	0.0036 (0.00027)
Age 85 Years and Older	0.0078 (0.00059)	0.0070 (0.00026)
Female (Male omitted)	-0.0019 (0.00026)	-0.0022 (0.00015)
African American (White omitted)	-0.000017 (0.00077)	-0.00095 (0.00035)
Hispanic	-0.00013 (0.00079)	-0.0014 (0.00026)
Asian	-0.00068 (0.0016)	-0.00065 (0.00035)
Native American	0.00030 (0.0045)	-0.000022 (0.0021)
Disease Stage 1.3-2.3 (Disease Stage 1.1 omitted)	0.00068 (0.00047)	0.00031 (0.00028)
Disease Stage 3.1-3.6	0.0084 (0.00047)	0.0079 (0.00033)
Disease Stage 3.7	0.014 (0.00045)	0.012 (0.00028)
Disease Stage 3.8	0.047 (0.0014)	0.046 (0.0013)
Emergency Room Admit (Routine Admit omitted)	0.0042 (0.0012)	0.0029 (0.00017)

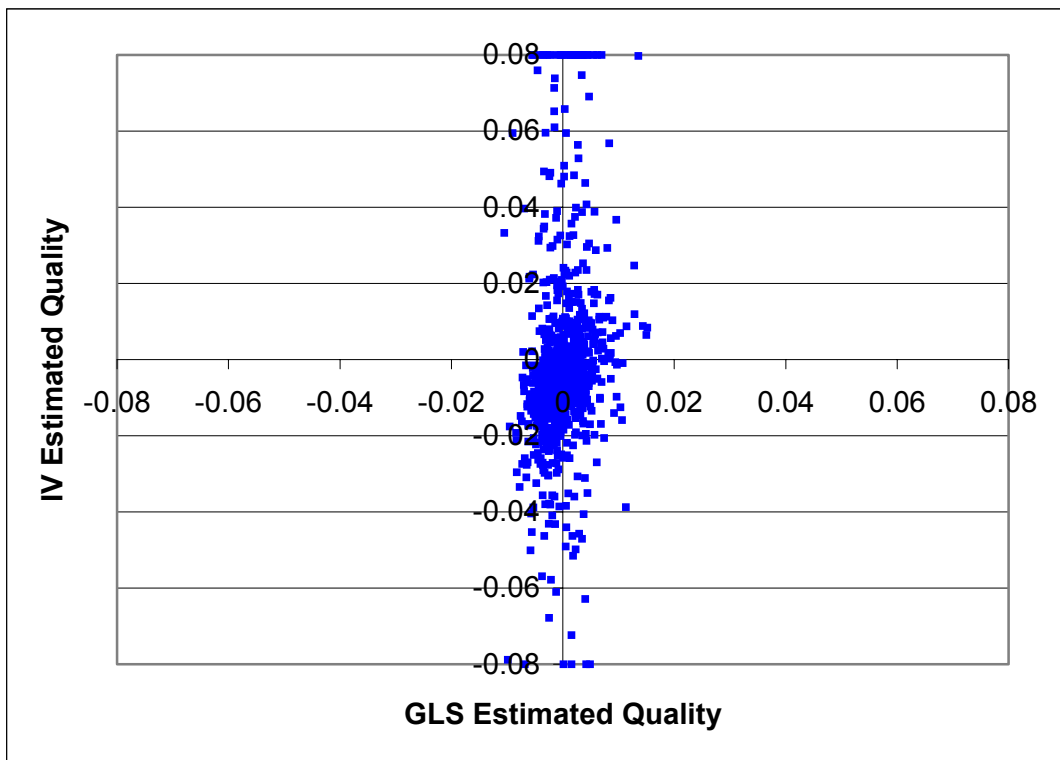
Note: regressions include hospital, year, and day since admission dummies.

Table 4
 Random-Effects Regression of Estimated Hospital Mortality
 on Hospital Characteristics: 1989-1994
 (Standard errors in parentheses)

Hospital Characteristic	Means and Standard Deviations	IV Estimated Mortality		GLS Estimated Mortality	
	(1)	(2)	(3)	(4)	(5)
Public	0.024 (0.15)	-0.0019 (0.0070)	-0.010 (0.011)	0.00017 (0.0014)	0.0048 (0.0022)
Private NFP (For-Profit omitted)	0.62 (0.49)	-0.0055 (0.0030)	-0.013 (0.0080)	-0.00016 (0.00047)	0.0015 (0.0010)
Public*Beds	827.0 (690.3)	—	7.11×10^{-6} (5.48×10^{-6})	—	-2.79×10^{-6} (9.71×10^{-7})
Private Not-For-Profit*Beds	293.5 (168.4)	—	2.28×10^{-7} (7.55×10^{-6})	—	-1.32×10^{-6} (2.13×10^{-6})
For-Profit*Beds	154.6 (62.6)	—	-0.00069 (0.00038)	—	5.87×10^{-6} (5.13×10^{-6})
COTH Member	0.071 (0.25)	—	0.0035 (0.0037)	—	0.00068 (0.00082)
Average LOS	9.73 (1.82)	—	-0.0024 (0.00077)	—	-0.00066 (0.00022)
Utilization Rate	0.53 (0.16)	—	-0.017 (0.0093)	—	-0.0012 (0.0014)
Long term Care Unit	0.17 (0.37)	—	-0.0037 (0.0024)	—	0.00032 (0.00055)
Geriatric Care Unit	0.094 (0.29)	—	-0.0043 (0.0046)	—	-0.00028 (0.00059)
Constant	—	0.0035 (0.0027)	0.045 (0.011)	0.00010 (0.00038)	0.0056 (0.0022)
N	844	844	844	844	844

Note: Standard errors are robust and are based on the White's methodology. "Average LOS" is mean length of stay for all Medicare pneumonia patients at the hospital. "COTH Member" is a dummy indicating membership in the Council of Teaching Hospitals. "Utilization Rate" is the number of patient-days divided by bed-days in a year. Regressions include annual dummies.

Figure 1
Scatterplot of the GLS and IV estimates of hospital quality
from separate years regression



Note: Observations that lie outside of the boundaries of the graph are plotted on the border.