Preliminary and incomplete
Do not quote

Estimating treatment effects with observational data: A new approach using hospital-level variation in treatment intensity

Mark McClellan
Stanford University and the NBER

Douglas Staiger
Dartmouth College and the NBER

March, 1999
(Revised July, 1999)

Jeffrey Geppert, Haruko Noguchi, and Yu-Chu Shen provided outstanding assistance with data management and analysis. Participants in seminars at various universities have provided helpful comments. We thank the Health Care Financing Administration (through California Medical Review, Inc.), the National Institute on Aging, and the Chiang Ching-kuo Foundation for International Scholarly Exchange for financial support

1. Introduction

Estimating the effect of medical treatments on patient outcomes is one of the central problems in medical research. Traditionally, randomized controlled trials have been the only definitive method for establishing such treatment effects. However, in response to a variety of perceived problems with randomized controlled trials, ranging from their expense to their external validity, there has recently been an increased interest in estimating treatment effects using observational data. The central problem with observational data, however, is that treatment is not randomly assigned across patients and is likely to be related to unmeasured patient characteristics and other factors which also influence outcomes.

Two basic approaches have been taken to this endogeneity problem (see McClellan and Noguchi, 1998). One approach has been to collect detailed patient information (e.g. from charts) and control directly for risk factors other than treatment that may influence patient outcomes. This approach is inevitably limited because it is difficult to measure all of the patient preferences, comorbidities, and other characteristics that may influence both treatment choice and outcomes. A second approach to the endogeneity problem has been to use instrumental variables (IV) estimation. In IV estimation one uses observed variables (instruments) that influence treatment decisions but are assumed to be unrelated to patient risk factors, to identify how treatment variation independently influences patient outcomes. This approach is limited by the availability of suitable instruments with adequate explanatory power, particularly in cases in which many treatments must be considered.

In this paper we propose a method of estimating treatment effects which combines these two prior approaches as well as makes more efficient use of the "good" variation in treatment use that appears unrelated to patient characteristics. We estimate patient-level outcome equations,

controlling for detailed patient risk factors, based on *hospital-level* variation in treatment intensity as an instrument for the treatment variables. After controlling for the observable risk factors, we argue that variation in treatment across hospitals is unlikely to be related substantially to unobserved risk factors affecting patient outcomes - unlike the variation in patient treatments within hospitals. In contrast to previous IV approaches, the large variation in practices across hospitals provides suitable instruments for evaluating multiple treatments of interest independently. Recent work on weak instruments suggest that standard IV estimation methods may yield biased estimates of treatment effects in our approach, because individual hospitals provide relatively weak instruments for treatment choice (Staiger and Stock, 1997). Therefore, we develop an alternative estimator, based on a General Method of Moments (GMM) approach closely related to the methodology developed in McClellan and Staiger (1997,1999), that does not suffer from this bias.

We apply our GMM method to estimating the effects of treatments provided by hospitals on survival following a heart attack. We use detailed medical chart review data and linked Medicare administrative records from the Cooperative Cardiovascular Project on over 180,000 elderly heart attack patients from 1994-1995. Our analysis includes a variety of medical and surgical treatments thought to improve survival following a heart attack, including cardiac catheterization, angioplasty, bypass surgery, the use of aspirin, and the use of beta blockers. To evaluate the methodology, we compare our GMM estimates to alternative estimates based on OLS regression and instrumental variables methods.

We present a number of empirical results that provide support for the GMM method. First, we document that there is considerable systematic variation in treatment intensity across hospitals. For example, holding patient mix constant we estimate that the average propensity to

do cardiac catheterization varied considerably across hospitals with a standard deviation of 14 percentage points around a national average of rate of 49%. Second, we find that the GMM method, which uses this variation in treatment intensity across hospitals to estimate treatment effects, yields estimates of treatment effects that are quite similar to recent IV estimates (McClellan et al., 1994) but that are considerably more precise. In addition, the GMM estimates appear to be quite robust to controlling for detailed information on patient severity. In contrast, OLS estimates, which we argue are fundamentally biased, differ significantly from the GMM and IV estimates and are not robust.

Substantively, our results suggest that both surgical and drug treatments for heart attack patients have significant and substantial effects on mortality. For example, we find that providing aspirin during the hospitalization reduces 1-year mortality by 5 percentage points, while Beta blockers reduce mortality by over 10 percentage points. Surgical treatments may have even larger effects.

2. Background

A. Estimating treatment effects

Unbiased estimation of treatment effects is of major interest in many branches of applied economics and statistics, for use in guiding individual decisions involving treatment use as well as policy decisions that may influence treatment use. The gold-standard method for estimating treatment effects without bias is the randomized controlled trial, in which subjects are randomized to receive either the treatment of interest or a "control" treatment protocol that differs only in a well-defined treatment of interest. But randomized trials are often very costly, and individuals often do not want to be subject to randomization when expensive treatments and

important outcomes are at stake, and when they have some prior beliefs about expected benefits with each alternative. Moreover, it may be especially difficult to enroll a sufficient number of patients to estimate treatment effects precisely for particular clinical types.

These limitations of randomized studies have resulted in continued interest in the estimation of treatment effects with observational data. One set of observational-data techniques is based on *direct comparisons* of treated and non-treated individuals. These methods use regression techniques, case-control comparisons within similar subgroups, or related adjustments such as propensity scores to account for observable differences between treated and non-treated groups that may result directly in outcome differences. Bias concerns arise because these methods assume no substantial correlation between treatment choice and many unmeasured factors or "omitted variables," including unmeasured differences in the individuals selected into the treatment groups as well as unmeasured differences in other treatments and environmental conditions to which the groups are exposed.

Another set of observational-data methods, *instrumental-variables* methods, rely instead on comparisons between groups that differ in an observable way that is assumed to influence treatment choice but not to influence outcomes directly -- much like randomization in a clinical trial. The instrumental-variables methods, however, will be biased if the instrumental variable is correlated with unobserved individual or environmental factors that also affect outcomes. Moreover, changes in the value of instrumental variables generally affect treatment for only a subset of individuals, so that the consequences of any such correlation are multiplied. Finally, application of this method in practice is limited by the availability of suitable instruments, particularly in situations where there are many dimensions of treatment. Lingering concerns about the potential for bias, along with practical data limitations in terms of suitable instruments

and measures of patient severity, have limited the acceptance of observational methods for estimating treatment effects.

B. Evidence on treatment effects for heart attack patients

In this paper, we focus on treatment of heart attacks. Heart disease is the leading cause of death in the United States, and heart attacks or acute myocardial infarctions (AMIs) are directly or indirectly responsible for most of these deaths. An important set of intensive treatments for heart attack care begins with cardiac catheterization, a procedure that visualizes blood flow to the heart muscle through continuous radiologic pictures of the flow of dye injected into the coronary arteries. If this procedure detects substantial blockages, it may be followed by a "revascularization" procedure intended to improve blood flow to the heart. The two commonly-used types of procedures are angioplasty (PTCA, or percutaneous transluminal coronary angioplasty), which involves the use of a balloon at the end of a catheter to eliminate blockages, and bypass surgery (CABG, or coronary-artery bypass graft surgery), a major open-heart surgical procedure to bypass the areas of blockage. In addition, patients can be treated with a variety of drugs following their heart attack. These include aspirin and thrombolytics (which inhibit clotting, and thereby improve blood flow to the heart) and Beta blockers and ACE inhibitors (which reduce demand on the heart, thereby reducing required blood flow to the heart). All of these drug treatments have been shown to reduce AMI mortality significantly, though most of the studies did not use elderly AMI patients.

Despite the importance of heart attacks for population health and the importance of these intensive procedures for health care resource use, the procedures have been studied in only a few randomized clinical trials. Several trials were performed for bypass surgery in the early 1980's and on angioplasty in the following years; in general, these trials found limited mortality benefits

in a few subgroups of patients.[1] Nonetheless, the procedures have become much more widely used in heart attack patients, for several reasons. First, the equipment quality and technical skill of personnel involved in the procedures has improved substantially since the time of the trials, leading to much lower complication rates. Second, trials on many types of heart disease patients, such as women and the elderly, were regarded as too costly to justify additional studies given the previous trial results. Third, as experience accumulated, fewer and fewer patients were willing to be randomized for such an important decision as an intensive cardiac procedure. As with many other intensive medical technologies, these heart procedures are now used in a much broader range of patients than have been explicitly supported by randomized trials.

Consequently, these procedures have been studied frequently using observational methods. Studies based on direct comparisons of treated and non-treated patients have generally found that intensive cardiac procedures like bypass surgery were associated with significant and substantial mortality reductions in these additional patients, even after accounting for observable differences. For example, using the propensity-score method, Rosenbaum and Rubin (1984) estimate a large improvement in functional status and in survival for patients with heart disease undergoing bypass surgery. In contrast, observational studies using instrumental-variables methods (based on administrative data with little clinical detail) have found small mortality effects in patients undergoing procedures, and the effects appeared to be due at least in part to

[1]Trials of bypass surgery versus no intensive procedures included a VA study and a European trial. Trials of angioplasty included Erbel et al. (1986), Simmoons et al. (1989), TIMI Study Group (1989), and Zijlstra et al., 1993. Most of these studies focused on the immediate use of angioplasty, rather than its use at all during the episode of treatment for heart attack. Reflecting changes in expectations about treatment benefits, recent trials have focused on narrower questions about use of the intensive procedures, such as the timing of catheterization (e.g., Califf et al., 1991), the choice between angioplasty and bypass surgery, and the use of catheterization in very narrow subsets of patients (e.g., VANQWISH Study Group, 1998).

acute treatments other than catheterization (e.g., McClellan, McNeil, and Newhouse, 1994; McClellan and Newhouse, 1997). Thus, different observational estimation methods appear to give substantially different results. The source of the differences is not clear, although recent work that controls for detailed patient severity information supports the validity of the instrumental variable estimates (McClellan and Noguchi, 1998). Nevertheless, these kinds of inconsistencies have plagued observational studies of treatment effects, and limited their relevance for clinical practice and policies intended to influence it.

Randomized studies of drug therapies have been easier to perform, at least in nonelderly male patients. Thrombolytics have been shown to reduce mortality significantly compared to noninvasive treatments; whether thrombolytics lead to outcomes that are as good as primary (immediate) angioplasty after AMI is more controversial. Beta blockers, aspirin, and (for patients with heart failure after AMI) ACE inhibitors have all been shown to reduce mortality after AMI. Some studies, primarily observational, have suggested that routine use of calcium channel blockers may lead to worse outcomes. Routine use of lidocaine is associated with worse outcomes, and nitrate use is suspected to have limited benefits (at least for mortality) in AMI patients treated with these other "modern" therapies.

3. Methods

A. Overview and identification

The key problem with estimating treatment effects with observational data is that treatment is likely to be correlated with unmeasured patient characteristics that also influence outcomes. Two simple examples illustrate the fundamental nature of this problem. First, a patient may be considered a poor candidate for surgery if admitted with a severe heart attack

along with multiple pre-existing conditions (such as diabetes or stroke). If this patient information effects outcomes but is unobserved by the econometrician, then OLS estimates will be biased toward overstating the effect of treatment on outcomes because "sicker" patients will not receive treatment. As a second example, consider the case of an average patient admitted with a heart attack but who never stabilizes and dies within hours of admission. This patient will not receive many treatments (e.g. bypass) because of their short survival, leading to a classic problem of reverse causation. Again, OLS estimates will tend to overstate the effect of treatment, since only survivors receive treatment. This second example is particularly problematic because even controlling for complete information on patient severity at admission would not eliminate the bias in OLS estimates. As a result, any estimate of treatment effects that relies on patient-level variation in treatment will be biased.

Our approach uses hospital-level variation in treatment intensity as an instrument to identify the relationship between treatment and outcomes. In other words, we rely on between-hospital variation rather than within-hospital (patient-level) variation. This approach requires that treatment intensity is uncorrelated with systematic unobserved differences in patient mix at the hospital level, e.g. aggressive treatment hospitals do not attract patients that differ *in ways that are unobserved at admission*. Of course, hospitals may attract different types of patients based on their style of treatment, e.g. teaching hospitals may attract more severe cases. However, a patient's choice of hospital will be based on the information available at the time of admission. Therefore, it is, at least in principle, possible to control for systematic differences in case mix across hospitals provided one has sufficient information on the patient at the time of admission. Moreover, the acute nature of the AMI admission decision limits opportunities for selection. As

a result, it is possible to estimate treatment effects based on hospital-level variation in treatment combined with data on patient severity at the time of admission.

In practice, the central drawback of this approach is that it is difficult to estimate treatment intensity based on the relatively small patient samples observed at each hospital. As a result, much of the observed variation in treatment rates across hospitals is due to random factors rather than systematic differences across hospitals in treatment intensity. Therefore, we use an estimation method that explicitly addresses this issue of estimation error in observed treatment rates.

Estimation of the model is a two-stage process. The first stage consists of patient-level regressions for the outcome measure and for each treatment measure. These regressions estimate reduced-form models in which the outcome/treatment depends on hospital fixed effects and patient characteristics. The regressions generate hospital fixed effects for both outcome and treatments, controlling for the patient mix admitted to the hospital. In the second stage of the estimation, treatment effects are estimated based on the relationship between the outcome fixed effect and the treatment fixed effects, carefully accounting for the estimation error in the estimates from the first stage. The second stage provides estimates of the systematic variation in treatment across hospitals, as well as estimates of the effect of each treatment on the outcome of interest.

B. Setup and Notation

Suppose that the outcome (Y_{ij}) for patient i admitted to hospital j depends on a 1xK vector of treatments (T_{ij}) and a 1xL vector of patient characteristics measured at admission (X_{ij}) according to the following equation:

(1) $$Y_{ij} = T_{ij}\beta + X_{ij}\Pi + \alpha_j + \varepsilon_{ij} ,$$

The parameters of interest are β (a Kx1 vector of treatment effects) and Π (a Lx1 vector of covariate effects). We decompose the error into two parts: A hospital component (α_j), which as discussed above is assumed to be uncorrelated with T; and a patient component (ε_{ij}) which may be correlated with T. Patient characteristics at the time of admission (X) are assumed to be uncorrelated with ε_{ij}.

The expected level of treatment is assumed to depend on patient characteristics and also to depend on the hospital to which the patient was admitted, so that:

(2) $T_{ij} = X_{ij}\Gamma + \delta_j + \nu_{ij}$,

Where equation 2 is a multivariate regression where Γ is a LxK matrix of parameters, and δ_j is a 1xK vector of hospital fixed effects representing the treatment intensity at hospital j on each of the K treatments. Again, patient characteristics at the time of admission are assumed to be uncorrelated with ν_{ij}.

Treatment is endogenous in equation 1, but we can plug equation 2 into equation 1 to arrive at the reduced form equation for the outcome variable:

(3) $Y_{ij} = X_{ij}\Psi + \theta_j + \omega_{ij}$,

Where: $\Psi = \Gamma\beta + \Pi$; $\theta_j = \delta_j\beta + \alpha_j$; and $\omega_{ij} = \nu_{ij}\beta + \varepsilon_{ij}$.

C. Estimation

The parameters of equation 1 cannot be estimated consistently by OLS because the treatment variable is endogenous. However, we can estimate the parameters in the reduced form equations (2 and 3) by OLS with hospital fixed effects (since X is uncorrelated with both ν and ε). With consistent estimates of the parameters in the reduced form equations, one could estimate the treatment effects by conventional instrumental variable methods, e.g. by two stage

least squares, using hospital fixed effects in the first stage (equation 2) as the instruments for treatment. In our application, however, first stage estimates of the hospital fixed effect (δ_j) are based on small samples in each hospital, with over half of all hospitals admitting fewer than 30 heart attack patients. Thus, the assumption that the first-stage parameters are consistently estimated is problematic. In particular, we face a situation of having many weak instruments, and recent work has documented that parameter estimates and standard errors from traditional IV estimators such as 2SLS are biased in this situation (see Bekker, 1994; Bound, Jaeger and Baker, 1995; Staiger and Stock, 1997).

Therefore, we take a different approach to estimation, similar to that taken by Deaton (1985) in estimating cohort models from many years of cross-section data, and similar to that taken by McClellan and Staiger (1997, 1999) in estimating hospital quality from many years of patient-level data. Our approach is also closely related to the hierarchical bayes model proposed by Chamberlain and Imbens (1996) for IV estimation with many instruments. The estimation is a two-stage process. We describe each of these in turn below.

First Stage

In the first stage, we estimate the patient-level reduced form models for the outcome measure (equation 3) and each treatment measure (equation 2) by OLS. Under the assumptions stated in section 3(B), these regressions provide unbiased estimates of all of the parameters of interest, along with an unbiased estimate of the covariance matrix for these estimates. Let $M_j = \left\langle \hat{\theta}_j, \hat{\delta}_j \right\rangle$ be the 1x(K+1) vector of estimated fixed-effect parameters for hospital j, and let S_j be the OLS estimate of the (K+1)x(K+1) covariance matrix for these parameters. Assuming that equations 2 and 3 are estimated on a large sample of patients, the patient covariate

parameters (Γ,Ψ) are known asymptotically and, as a result, the estimates M_j are independent across hospitals.

Second Stage

In the second stage, we use the hospital-level data from the first stage (M_j,S_j) to estimate the remaining parameters by General Method of Moments (GMM). Note that the fixed effect estimates from the first stage are equal to the true hospital-specific intercepts plus estimation error:

(4) $$M_j = \left\langle \hat{\theta}_j, \hat{\delta}_j \right\rangle = \left\langle \theta_j, \delta_j \right\rangle + \zeta_j$$

Where $\left\langle \theta_j, \delta_j \right\rangle = \left\langle \delta_j \beta + \alpha_j, \delta_j \right\rangle$, and ζ_j is the 1x(K+1) vector of estimation error which is mean zero and uncorrelated with $\left\langle \theta_j, \delta_j \right\rangle$. This implies that:

(5) $$E\left(M_j' M_j\right) = E\left(\left\langle \theta_j, \delta_j \right\rangle' \left\langle \theta_j, \delta_j \right\rangle\right) + E\left(\zeta_j' \zeta_j\right)$$

Note that the first stage estimate of the covariance matrix for the fixed effect parameters (S_j) is an unbiased estimate of the variance in the estimation error, $E\left(\zeta_j' \zeta_j\right)$. Therefore, we have:

(6) $$E\left(M_j' M_j - S_j\right) = E\left(\left\langle \theta_j, \delta_j \right\rangle' \left\langle \theta_j, \delta_j \right\rangle\right)$$

Equation 6 states that the variance of the true fixed effect parameters is equal in expectation to the variance of the estimated fixed effect parameters minus the estimation error variance. Plugging in to equation 6 for θ_j and assuming that hospital treatment intensity (δ_j) is independent of hospital-level variation in unobserved case-mix (α_j) yields:

(7) $$E\left(M_j' M_j - S_j\right) = \begin{bmatrix} \beta' \Omega_\delta \beta + \sigma_\alpha^2 & \beta' \Omega_\delta \\ \Omega_\delta \beta & \Omega_\delta \end{bmatrix}$$

where $\Omega_\delta = Var(\delta_j)$, and $\sigma_\alpha^2 = Var(\alpha_j)$.

Equation 7 provides the basis for a just-identified GMM estimator of the treatment effect parameters (β), along with estimates of the variance (and covariance) of treatment intensities across providers (Ω_δ) and the variance of unexplained outcome differences across providers (σ_α^2). In particular, we estimate the parameters by setting the theoretical moments (the right hand side of equation 7) equal to an unbiased sample estimate of these moments calculated as the weighted average of $M_j'M_j - S_j$ across hospitals, using the number of admissions at each hospital as weights. Standard errors for the estimates are calculated as in Chamberlain (1984).

Our GMM method can be interpreted in a simple way. The fixed effect in the reduced form outcome equation is related to the fixed effect in the treatment equations by the equation:

(8) $\theta_j = \delta_j\beta + \alpha_j$

where α_j is independent of δ_j. Thus, if we observed θ_j and δ_j without any estimation error (or assumed that our estimates were consistent, so that the measurement error disappeared asymptotically), then equation 8 could be estimated by weighted least squares using hospital-level data and weighting by the number of patients in each hospital. Traditional two-stage least squares estimates of β, using hospital fixed effects as instruments, estimate equation 8 exactly in this way, replacing the unobserved parameters θ_j and δ_j with their estimates from equations 2 and 3. Our method simply estimates equation 8 correcting for the (correlated) measurement error in the estimates $\hat{\theta}_j$ and $\hat{\delta}_j$. Thus, rather than estimating β with the usual least squares formula (ignoring the weights) of $\left(\hat{\delta}'\hat{\delta}\right)^{-1}\left(\hat{\delta}'\hat{\theta}\right)$, we use estimates of the moment matrices that correct for the measurement error. For example, $\frac{1}{N}\left(\hat{\delta}'\hat{\delta}\right)$ overstates the variance in treatment across

hospitals ($E(\delta'\delta) \equiv \Omega_\delta$) because of the estimation error in $\hat{\delta}$. Therefore, our method "subtracts off" the estimation error and estimates the variance in treatment across hospitals (Ω_δ) with $\frac{1}{N}(\hat{\delta}'\hat{\delta}) - \overline{S}_{\hat{\delta}}$, where $\overline{S}_{\hat{\delta}}$ is the average estimation error variance for the treatment fixed effects ($\hat{\delta}_j$).

Our key identifying assumption is that hospital treatment intensity (δ_j) is uncorrelated with any unobserved factor influencing the hospital's average outcome (α_j). As argued earlier, as long as we can condition on the same information about patient severity that was used by patients in selecting hospitals, there is no reason to expect that treatment intensity is correlated with unobserved casemix differences across hospitals. However, treatment intensity may still be correlated with other unobserved hospital factors that influence outcomes. For example, any omitted aspect of treatment at a hospital that influences outcomes and is correlated with the included treatments will violate our identifying assumption and bias the results. Thus, as with any production function estimate, the estimated treatment effect may be the result of related treatments that were not included in the model. We discuss further evidence on the validity of this key assumption below.

4. Data

To explore the utility of the GMM method versus other observational estimation methods, we use data from the Cooperative Cardiovascular Project (CCP), a major policy initiative to improve the quality of care for Medicare beneficiaries with AMI undertaken by the Health Care Financing Administration (HCFA). During the "national" phase of the project, HCFA conducted standardized abstractions of the medical records of all Medicare beneficiaries

hospitalized with a reported AMI over an eight-month period at essentially all hospitals in the United States that had not participated in a four-state "pilot" phase. The eight-month sampling frame was continuous at each hospital, and all sampling occurred between April 1994 and July 1995. Marciniak et al. (1998) provides more details on CCP goals, sampling and data collection strategy, and methods used to assure standardization and completeness of the medical record reviews. Altogether, charts were abstracted for approximately 180,000 AMI patients. The sample we use includes all patients admitted with an AMI to a hospital that had at least 3 such admissions in this time frame, with patients assigned to the initial hospital at which they were admitted. These data were linked to Medicare administrative records (enrollment and hospitalization files), which have been used in previous observational studies of AMI practices and outcomes but do not include the clinical details present in the medical record abstracts. The enrollment files include comprehensive all-cause mortality information from Social Security records.

The CCP data provides extensive clinical information on treatments and outcomes for each patient, along with a detailed set of patient covariates covering demographic information, information on the presence of comorbidities, and information on the severity of the heart attack on admission. In our analysis, we use the following variables:

Treatment measures: From administrative claim data, we calculate 90-day treatment rates for cardiac catheterization, PTCA (angioplasty), and cardiac bypass (CABG) surgery, based on whether the patient received the treatment within 90 days of the initial heart attack. With the CCP data, we also calculate whether the patient received aspirin or beta blockers during the hospital stay.

Outcome measures: We use death dates validated by the Social Security administration to calculate survival times from the initial hospitalization. Based on these dates, we compute 1-day, 30-day, and 1-year mortality rates.

Patient Covariates: The CCP data include detailed measures of patient clinical characteristics. These measures enable us to control for important clinical information observed at admission that is likely to influence patient outcomes. We include three types of patient variables:

(1) demographic measures including gender, race, age, and urban residence;

(2) comorbidity measures including measures of mobility, dementia, diabetes, CVA/Stroke, Angina, and CHF or pulmonary edema; and

(3) severity measures including Killip Class, heart rate, mean arterial pressure, time since chest pain began, blood urea nitrogen, and whether the patient is verbally oriented.

For detail on how these variables are constructed, see McClellan and Noguchi (1998); for a more detailed description of comorbidity and severity variables that are predictive of mortality in CCP, see Normand et al. (1997).

Table 1 provides summary statistics for the key treatment and outcome variables in our analysis. We report means and standard deviations for the hospital-level treatment and outcome measures (e.g. fixed effects from equation 2 and 3) adjusted for the detailed list of patient covariates just described. All statistics have been weighted by the number of AMI admissions at each hospital. Mortality for heart attack patients is relatively high, with 5.7% mortality in the first day, 18.9% in the first 30 days, and 32.3% in the first year following the heart attack. Estimated mortality rates also vary considerably across hospitals, with a standard deviation of 30-day mortality across hospitals of 7.1 percentage points. Of course, some of this variation is due to

estimation error in each hospital's estimate, as the average hospital admits only 45 patients. Just under 50% of patients receive catheterization, with roughly 1/3 of these patients going on to have angioplasty (18.8%), and another 1/3 going on to have bypass (15.5%). Treatment rates for the medications we study, beta blockers (44.3%) and aspirin (77.2%), are considerably higher – though substantially lower than most experts believe is appropriate. The variation in these estimates of treatment rates across hospitals is large, with a standard deviation across hospitals ranging from 7.6 (bypass) to over 15 percentage points (catheterization, beta blockers).

Finally, following McClellan et al. (1994), we construct a variable ("differential distance") for each patient measuring the difference between the distance to the nearest hospital doing catheterization, and the distance to the nearest hospital not doing catheterization. Distance estimates are measured between population centroids of the zip code of residence and the zip code of each hospital. This variable serves as an instrument for catheterization in that it is correlated with receiving catheterization (patients nearer a catheterization hospital are more likely to receive this treatment), and is arguably unrelated to unobserved patient severity. For a more details on the construction and justification of this variable, see McClellan et al. (1994) and McClellan and Noguchi (1998).

5. Results

In this section we report our GMM estimates for a variety of models. To explore the robustness of these methods, we report estimates based on various combinations of outcomes, treatments, patient covariates, and estimation methods. We begin by reporting our GMM estimates of the variance in treatment intensity and in outcomes across hospitals, along with the correlation in these treatment and outcome measures across hospitals. These estimates are of

independent interest, since they document the extent of practice and outcome variations across hospitals after accounting for variation in patient mix. We then report estimates of simple models in which the only treatment variable is whether the patient received cardiac catheterization within 90 days of the initial heart attack. We compare our GMM estimates to simple OLS estimates of the treatment effect and, following McClellan et al. (1994) and McClellan and Noguchi (1998), to IV estimates using differential distance to a catheterization hospital as an apparentlly valid instrument for treatment (see the data section for a discussion of how this variable was created). Finally, we report estimates from models that include additional treatment variables, including whether a patient had bypass surgery or angioplasty within 90 days of admission, or received aspirin or beta blockers during their hospital stay.

Estimates of Variance and Correlation in Treatment and Outcomes Across Hospitals

How much variation is there in casemix-adjusted mortality and treatment rates across hospitals? Our estimates suggest that there is substantial variation in both treatment rates and mortality rates across hospitals. In Tables 2 and 3 we report the GMM estimates for the variation across hospitals in measures of mortality and treatment intensity. Along the diagonal of each table, we report estimates of the standard deviation in each measure, while the off-diagonal elements of each table report the estimated correlation between the measures. All of these estimates are quite precise, with standard errors for each estimate given in parentheses. The estimates reported in these tables are based on mortality and treatment rates that were adjusted for the detailed list of patient covariates available in the CCP data.

Table 2 reports estimates for our outcome measures: 1-day mortality, 30-day mortality and 1-year mortality. The estimated standard deviation in mortality rates across hospitals is substantial, ranging from 2.2 percentage points for 1-day mortality (relative to a base 1-day

mortality rate of 5.7%) to 3.4 percentage points for 1-year mortality (relative to a base of 32.3%). Note that these GMM estimates are estimates of the true variation in mortality rates across hospitals. Thus, they are smaller than the standard deviations reported in Table 1 because they have corrected for the over-dispersion in the estimated mortality rates across hospitals due to estimation error. The variation in mortality rates across hospitals is also highly correlated across the mortality measures, with correlation ranging from 0.6 to over 0.9. Thus, differences in mortality across hospitals that appear in short-term mortality measures persist through 1-year mortality.

Table 3 reports similar estimates for our five treatment measures. Again the standard deviation in treatment rates across hospitals is quite large for most of the measures, ranging from 4.8 percentage points for bypass surgery to 13.6 percentage points for catheterization. Thus, treatment patterns clearly differ across hospitals, even for inexpensive treatments such as aspirin (SD=9.5 percentage points). Treatment rates are positively correlated, particularly among the surgical treatments and among the drug treatments, suggesting that these treatments are complements: hospitals that use one surgical treatment aggressively tend to be aggressive on other surgical treatments as well; bypass surgery is not a substitute in practice for the angioplasty, the other revascularization procedure. Not surprisingly, rates of catheterization are very highly correlated with rates of both bypass (correlation of 0.7) and angioplasty (correlation of 0.9) across hospitals. As a result, it will be difficult to distinguish the effects of catheterization from bypass and angioplasty in our estimates based on variation across hospitals. Therefore, we focus our attention on models that include either catheterization or bypass/angioplasty (revascularization) and acknowledge, for example, that the estimated treatment effect for

catheterization is likely to result primarily from this variable proxying for bypass and angioplasty, rather than any direct benefit of catheterization itself.

Estimates of Treatment Effects for Catheterization

Table 4 reports estimates of the treatment effect (β) for a simple version of equation 1 in which the dependent variable is mortality within a specified time period after admission and the only treatment variable is whether the person received cardiac catheterization within 90 days of their heart attack. Each column reports estimates for a selected mortality window, ranging from 1-day (column 1) to 1-year (column 3). The left panel reports estimates that control for the full set of patient covariates available in the CCP data. The right panel reports estimates that only control for a limited set of demographic variables (gender, race, age, urban residence) commonly available in claims data. Finally, within each panel, the first row reports our GMM estimates, the second row reports IV estimates based on differential distance to a hospital performing catheterization, and the third row reports OLS estimates.

If the assumptions required for consistency of the GMM estimator are correct, then the IV estimates should also be consistent but less efficient because they use less of the variation in treatment rates across hospitals. In contrast, the OLS estimates will most likely be biased toward overstating the treatment effect due to reverse causation (patients who die soon after admission will not receive catheterization). This is precisely the pattern observed in the estimates that control for the full list of patient covariates (the left hand panel). The GMM point estimates are not significantly different from the IV estimates, but are considerably more precise with standard errors 3-4 times smaller than the IV estimates. At the same time, the GMM estimates of the treatment effect are significantly smaller than OLS estimates. For example, in column 2 we estimate that catheterization is associated with a reduction in 30-day mortality of 8.0 percentage

points with a standard error of 1.0 percentage point. Instrumental variable estimates of the treatment effect are slightly larger (11.3 percentage points) but not significantly different from the GMM estimates because of their large standard error (3.4 percentage points). Finally, the OLS estimate of the treatment effect is 12.8 percentage points, more than 50% larger than the GMM estimate. Because of the precision of the OLS estimate, we can easily reject that OLS and GMM estimates are equal.

The right panel of Table 4 reports a similar set of estimates to the left panel, but for models that control for a more limited set of demographic variables. These estimates inform us about the robustness of the GMM, IV and OLS methods when using the type of data commonly found in claims databases. The results demonstrate that, at least in this application to heart attack mortality, the GMM and IV methods are quite insensitive to the exclusion of detailed patient covariates, especially at acute periods after AMI. For example, the GMM and IV estimates of the effect of cardiac catheterization on 30-day mortality are nearly identical to the estimates in the left panel of Table 4 that controlled for a more detailed set of covariates. In contrast, OLS estimates, which rely on the patient-level variation in treatment, are quite sensitive to the exclusion of detailed patient covariates, with the estimated treatment effect nearly doubling when we use the more limited set of control variables.

Overall, the results in Table 4 support four important conclusions. First, the GMM estimates are quite similar to, but more precise than, estimates from the one-treatment IV model of McClellan et al. (1994). At the same time, the GMM method may be more easily used in practice. It does not require one to identify instruments specific to each application (such as the differential distance measure used in McClellan et al.), and thus can be widely used for treatment analysis provided that patient sorting to hospitals based on unobserved severity is not a

significant problem. Thus our method substantially expands the IVs available for estimating treatment effects with observational data. A second conclusion is that OLS estimates appear to be biased toward overstating the treatment effect, even when the regressions control for the extensive patient covariates available in the CCP dataset. Thus, OLS appears to be fundamentally flawed as a practical method for estimating treatment effects. A third conclusion is that the GMM method, like the IV method of McClellan et al., appears to provide unbiased estimates even in datasets with limited patient covariates. Thus, the GMM method provides a practical method for estimating treatment effects using commonly available claims datasets. Finally, the GMM estimates suggest that catheterization is associated with modest improvement in patient survival of about 5 percentage points at 1 day, 8 percentage points at 30 days, and 12 percentage points by 1 year. Obviously, these effects are not the result of catheterization per se, but rather the result of other treatments that are correlated with catheterization. Indeed, McClellan et al. (1994) and McClellan and Noguchi (1998) found that most if not all of the apparent effect of catheterization in IV analysis could be explained by its correlation with other treatments. Thus, we now turn to estimating models that control for a more detailed list of treatments.

Estimates of Treatment Effects in More General Models

In this section we compare GMM and OLS estimates of models that include surgical (bypass, angioplasty) and drug (aspirin, beta blockers) treatments in addition to catheterization. One of the great advantages of the GMM method is that it more easily extends to models with many treatments. In contrast, while IV estimates as in McClellan et al. (1994) have proven feasible for analyzing several treatments (e.g., McClellan and Noguchi, 1998), they are difficult

to apply to models with many treatments because one must identify additional instruments for each treatment and these instruments must generate independent variation in each treatment.

Table 5 reports treatment effect estimates for models in which the dependent variable is again mortality, but now we include three treatment variables: catheterization, receiving aspirin in the hospital, and receiving beta blockers in the hospital. Thus, catheterization serves as a proxy for invasive treatment, while aspirin and beta blockers capture some of the key drug treatments that are believed to improve patient survival. The layout of Table 5 is similar to that of Table 4. The top panel reports GMM estimates of the coefficients for each treatment, while the bottom panel reports OLS estimates of the coefficients. The left panels report estimates that control for a full set of patient covariates, while the right panels report estimates that control for a limited set of demographic variables. Finally, within each panel, each column reports estimates using either 1-day, 30-day or 1-year mortality as the dependent variable.

Focussing first on the GMM estimates, we find that controlling for aspirin and beta blockers reduces the estimated effect of catheterization substantially. Based on the models with a full set of patient covariates, we estimate effects of catheterization on mortality that are about one third as large (as compared to Table 4) at 1-day and 30-days, and about 20% smaller at 1 year. Aspirin is estimated to have substantial effects on short-term mortality, peaking at a 7.2 percentage point reduction in 30-day mortality. In contrast, the effect of Beta blockers (like catheterization) appears to cumulate so that by 1 year after the heart attack they have reduced mortality by 10.0 percentage points. All of these estimates are fairly precise, with standard errors of 0.7 to 2.5 percentage points.

Based on clinical knowledge, one might not expect the effect of short-lived hospital treatments (such as Beta blockers) to have effect on mortality that was significantly larger at 1

year than at 30 days. Given that many important treatments are still omitted from this model, it is likely that some of this effect is coming from other omitted treatments (such as use of medications after leaving the hospital) that are correlated with receiving Beta blockers in the hospital. Thus, until models are estimated with a more complete list of treatments, we believe that these preliminary results should be interpreted cautiously.

As in Table 4, OLS estimates larger effects of catheterization. In fact, controlling for aspirin and beta blockers has little effect on OLS estimates of the effect of catheterization on mortality. The OLS estimates of the effect of aspirin are also larger than the GMM estimates. In contrast, OLS estimates of the effect of beta blockers on mortality are smaller than the GMM estimates. As argued above, OLS estimates will be biased by the fact that patients who die quickly will be less likely to receive some treatments, even if one controls perfectly for patient severity at the time of admission. Therefore, we believe that these differences between the OLS and GMM estimates reflect substantial bias in the OLS estimates resulting from treatment variations within hospitals; the GMM estimates remove such bias.

Once again, the GMM estimates are similar when we control for a more limited set of patient covariates. In particular, both the estimated effects of catheterization and of beta blockers are little changed. Only the aspirin effects appear to change systematically, increasing by 3-5 percentage points when we use the more limited patient covariates. The effect of catheterization on mortality also appears slightly larger at 1 year. In contrast, the OLS estimates are quite sensitive to changes in the patient covariates, with estimated effects of each treatment on mortality nearly doubling by 1 year when we control for a more limited set of covariates.

Catheterization, of course, is a diagnostic procedure and not a treatment per se. Thus, including it as a treatment in these regressions is somewhat ad hoc, in that it is clearly being

included as a proxy for invasive treatment. Alternatively, we could drop catheterization as a treatment variable and replace it with direct measures of invasive treatments. In Table 6, we report such estimates.[2] The estimates are for similar specifications to those reported in Table 5, but replacing catheterization with two surgical treatments: did the patient receive bypass surgery within 90 days of their heart attack, and did the patient receive angioplasty within 90 days of their heart attack. These models continue to include aspirin and beta blockers as treatment variables.

The results are quite striking, with estimated effects of bypass in particular being too large to be clinically plausible. Based on the GMM estimates controlling for a detailed set of patient coviates, both surgical treatments are found to have significant effects on 1-year mortality. The effects of angioplasty are relatively small and insignificant on short-term mortality, but angioplasty is estimated to reduce 1-year mortality by 7.4 percentage points. Bypass surgery is estimated to have significant effects on all mortality measures, with bypass estimated to reduce mortality at one year by over 22 percentage points. These results suggest that the overall reduction in mortality associated with catheterization (about 10 percentage points from Table 5) can be decomposed into two effects. About one third of the patients receiving catheterization go on to have bypass surgery, and their mortality is estimated to be reduced by 22 percentage points. Another third of the patients receiving catheterization go on to have angioplasty, and their mortality is estimated to be reduced by 7 percentage points.

[2] When we estimate models that include catheterization as well, the coefficients on bypass and angioplasty change very little, and the coefficient on catheterization is small in magnitude and insignificant. However, standard errors on all of these coefficients increase by a factor of 2-4 because of the strong correlation in these treatments.

Why are the estimated effects of bypass so large? The most likely reason is that, as was the case with catheterization, hospitals that perform revascularization most extensively are among the most technologically sophisticated hospitals. Many other treatments may also be used more extensively by these hospitals. As a result, the large estimated effect of bypass may be the result of bypass proxying for a set of treatments that have not been directly controlled for in the regression. In other words, even if the GMM method eliminates bias resulting from treatment depending on patient severity, hospitals with high bypass rates may still have much lower mortality rates because of other treatments they provide rather than the direct effect of bypass. In future work, we intend to explore the full potential of the GMM methods to control for these likely correlated treatments.

The remaining GMM estimates in Table 6 tell a similar story to the estimates from earlier tables. The estimated effects of aspirin and Beta blockers does not change much when we include bypass and angioplasty in the model. Once again, the GMM estimates change little when we control for demographic variables only. As in earlier specifications, the OLS estimates are quite different from the GMM estimates and much more sensitive to changes in the set of patient covariates. The most striking difference is that OLS estimates that bypass reduces 1-year mortality by 10.7 percentage points, compared to the GMM estimate of 22.2 percentage points.

We assume that the patient-level variation in treatment used by OLS is endogenous, and therefore the OLS estimate is biased, because unexpected mortality (or survival) directly effects treatment decisions. One way of checking this assumption is to try including a treatment that on *a priori* grounds should have little effect on mortality. Catheterization is a good example, as this is a diagnostic treatment that should have little direct effect on mortality. When we add catheterization as a treatment, the GMM method estimates that catheterization has a small and

insignificant effect on mortality, as expected. Moreover, adding catheterization to the models reported in Table 6 has little effect on the GMM estimates of the other treatment effects, although the standard errors of the estimates increase substantially due the high correlation among these treatments. For example, the estimated percentage point reduction in 1-year mortality is 10.4 for angioplasty (SE=7.0), 25.4 for bypass (SE=9.8) and –2.7 for catheterization (SE=6.3). In contrast, when we include catheterization in the OLS models, the estimated effect of bypass and angioplasty fall even further (to around 5 percentage points on 1-year mortality), while the estimated effect of catheterization is large and significant (11.8, SE=0.3 for 1-year mortality). Since we would not expect catheterization to have such large effects, this suggests that the OLS estimates are indeed biased by the endogeneity of catheterization use within hospitals.

Overall, the results of this section provide further support for the GMM method. We are able to estimate fairly precise treatment effects in models that control for multiple dimensions of treatment. These estimates are quite similar whether we control for demographic variables only, or for a detailed list of patient comorbidity and severity measures available in the CCP data. Finally, the estimated treatment effects of drugs are broadly consistent with medical knowledge. Aspirin is found to have substantial immediate effects on short-term mortality, while beta blockers have more gradual, but still substantial, effects on mortality. While catheterization *per se* is found to have no significant effect on mortality, invasive treatments following catheterization do appear to have effects on long-term mortality. However, given the high level of correlation among the treatments we have included in this preliminary analysis, we expect that at least some of these treatment effects are the result of other related treatments being provided to patients but not yet incorporated into our analysis.

5. Conclusion

Estimating the effects of medical and surgical treatments in a valid manner is the foundation of evidence-based medicine. Despite advances in our ability to collect and analyze large databases of patient claims, there remains no general method for estimating valid treatment effects for the large number of treatments that can vary in observational data.

In this paper, we propose such a method and demonstrate the method's usefulness in estimating the effect of medical and surgical treatments on mortality among elderly heart attack patients. The method relies on hospital-level variation in treatment intensity, along with sufficient data on patients to control for relevant medical information these patients used in selecting a hospital. Thus, it eliminates bias arising from treatment selection based on unobservable patient characteristics within hospitals – which is plausibly the major source of selection bias in observational studies of acute conditions like AMI. In our application, this method appears to produce precise estimates of treatment effects that are largely robust to whether or not one controls for detailed patient-level severity measures.

On a substantive level, we find that the effects of drug treatments (aspirin, beta blockers) and surgical treatments (bypass, angioplasty) on survival are both significant and substantial. We emphasize that our preliminary results here only begin to use the full potential of our methods; in ongoing research, we are expanding our models to accommodate more of the many treatments that can and do vary for AMI patients. These more complete models will enable us to determine how much of the estimated treatment effects in our preliminary results, which remove selection bias arising from patient-specific treatment decisions within hospitals, are the result of other treatments whose use we have not yet modeled. We will also explore further whether a small amount of residual across-hospital selection bias remains.

In addition, there remains substantial variation across hospitals in the rate at which heart attack patients are provided with these treatments. Understanding the reasons for this variation in treatment rates is an important topic of future research. Also, since these treatments vary dramatically in their costs as well as estimated effects on survival, it will be important for future work to evaluate the cost-effectiveness of these alternative treatments.

Finally, we believe that the method we propose is generally applicable, and can be used in many applications to estimate treatment effects using the type of observational data commonly available in claims databases. The method can, in principal, be applied to any outcome and any combination of treatments, as long as there is variation in treatment rates across hospitals and as long as hospital-level selection bias is not substantial. More generally, the method could be applied wherever there is systematic variation at some aggregate level, such as physician, medical group, or geographic region. Of course, there may be particular features of our application to elderly heart attack patients that led to the apparent success of the method. Evaluating the performance of this method in other applications is an important area for future work.

References

Bekker, P. A., 1994, "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 1994, 657-681.

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, 1995, 443-450.

Califf, R.M. and the TAMI V-A Study Group, 1991, "Evaluation of Combination Thrombolytic Therapy and Timing of Cardiac Catheterization in Acute Myocardial Infarction," *Circulation* 83: 1543-1556.

Chamberlain, G., 1984, "Panel Data," in X. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics, Vol. 2*, New York, Elsevier Science, 1984, 1247-1318.

Chamberlain, G. and G. Imbens, 1996, "Hierarchical Bayes Models with Many Instrumental Variables," Harvard Institute of Economic Research, Paper Number 1781.

Deaton, Angus, 1985, "Panel Data from Time Series of Cross-Sections," *Journal of Econometrics*, 30(1-2), Oct.-Nov. 1985, 109-26.

Erbel, R., T. Pop, K.J. Henrichs, K. von Olshausen, C.J. Schuster, H.J. Rupprecht, C. Steuernegel, and J. Meyer, 1986, "Percutaneous Transluminal Coronary Angioplasty after Thrombolytic Therapy: A Prospective Controlled Randomized Trial," *Journal of the American College of Cardiology* 8: 485-495.

Marciniak, Thomas, Ellerbeck, Edward, Radford, M.J., et al., 1998, "Improving the Quality of Care for Medicare Beneficiaries with Acute Myocardial Infarction," *Journal of the American Medical Association*, 1998, forthcoming.

McClellan, M., B.J. McNeil, and J.P. Newhouse, 1994, "Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality? Analysis Using Instrumental Variables," *Journal of the American Medical Association*, 272(11),1994, 859-866.

McClellan, Mark and Haruko Noguchi, 1998, "Valid Interpretation of Treatment Effect Estimates Using Observational Data: Treatment of Heart Attacks in the Elderly," working paper, February,1998.

McClellan, M. and J. P. Newhouse, 1997, "The Marginal Cost-Effectiveness of Medical Technology: A Penel Instrumental-Variables Approach," *Journal of Econometrics*, 77(1), March 1997, 39-64.

McClellan, M. and D. Staiger, 1997, “Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals,” in D. Cutler (ed.) *Not-for-Profit Hospitals*, Chicago, University of Chicago Press, forthcoming.

McClellan, M. and D. Staiger, 1999, “The Quality of Health Care Providers,” working paper, 1999.

Rosenbaum, Paul, and Rubin, Donald, 1984, “Reducing Bias in Observational Studies Using Subclassification on the Propensity Score,” *Journal of the American Statistical Association* 79: 516-524.

Simmoons, M.L., A.E.R. Arnold, A. Betriu, D.P. DeBono, J. Col, F.C. Dougherty, R. Von Essen, H. Lambertz, J. Lubsen, B. Meier, P.L. Michel, P. Raynaud, W. Rutsch, G.A. Sanz, W. Schmidt, P.W. Serruys, C. Thery, R. Uebis, A. Vahanian, F. Van de Werf, G.M. Willems, D. Wood, and M. Verstaete for the European Cooperative Study Group for Recombinant Tissue-Type Plasminogen Activator (rTPA), 1988, "Thrombolysis with Tissue Plasminogen Activator in Acute Myocardial Infarction: No Additional Benefit from Immedicate Percutaneous Angioplasty," *Lancet* 197-202.

Staiger, Douglas and James H. Stock, 1997, “Instrumental Variables Regression With Weak Instruments,” *Econometrica*, 65(3), May 1997, 557-586.

Thrombolysis in Myocardial Infarction (TIMI) Study Group, 1989, "Comparison of Invasive and Conservative Strategies after Treatment with Intravenous Tissue Plasminogen Activator in Acute Myocardial Infarction," *New England Journal of Medicine* 320: 618-627.

VANQWISH Study Group, 1998

Zijlstra, Felix, et al., 1993, "A Comparison of Immediate Coronary Angioplasty with Intravenous Streptokinase in Acute Myocardial Infarction," *New England Journal of Medicine* 328:680-684.