

**How Do Good Hospitals Do It?
Estimating the Effects of Medical Practices**

Jeffrey Geppert
NBER

Mark McClellan
Stanford University and the NBER

Douglas Staiger
Dartmouth College and the NBER

October, 2000

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 (through NBER), and the Chiang Ching-kuo Foundation for International Scholarly Exchange for financial support.

1. Introduction

In this paper, we develop and apply methods for addressing two core problems in evaluating the quality of medical care: determining the effect of medical treatments on patient outcomes, and determining whether and why health outcomes differ across health care providers.

Randomized controlled trials are generally regarded as the only definitive method for estimating treatment effects. However, in response to a variety of perceived problems with randomized controlled trials, ranging from their expense and timeliness to their external validity and the “ethical” problems of conducting trials for many commonly-used treatments, some investigators have advocated supplementing clinical-trial results with analyses of various types of observational data. As medical data systems improve, increasingly detailed observational data on large patient populations have become available for potential use in such studies. These data innovations do not eliminate the central problem with observational data: treatments are not randomly assigned across patients and are likely to be related to hard-to-measure factors that also influence outcomes.

Two major approaches have been taken to this problem of endogeneity in estimating treatment effects (see McClellan and Noguchi, 2000). 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; relative to the many complex patient characteristics that are observed by patients and their health care providers and that may influence treatment choices, the effects of particular medical treatments are often quite modest. A second approach 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 (see McClellan and Newhouse, 2000).

In this paper we propose a method of estimating treatment effects that combines these two prior approaches, that makes more efficient use of “good” variation in treatments that appears unrelated to individual characteristics, and that clearly describes which treatments are separately identified. We estimate models of the effects of medical treatments on individual patients based on *hospital-level* and *area-level* variation in treatment intensity as an instrument for the treatment variables. We focus on hospital-level analysis; 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 – in contrast to the variation in patient treatments within hospitals. To do this, we develop a new estimator, based on a General Method of Moments (GMM) approach derived from methods developed in McClellan and Staiger (1999a, 1999b, 2000), which allows more precise identification of true or systematic variation in treatments and outcomes across hospitals. Because these treatment variations occur at the level of hospitals or small areas and do not appear to be related significantly to differences in patient mix, they provide a much more powerful approach to identifying consequences for patient outcomes.

This feature of our approach also makes it relevant to evaluating quality of care of different health care providers. Because patients can never be effectively randomized to alternative hospitals or other medical providers, observational data will always provide the foundation for comparisons of quality across providers. In addition to the problems of bias that generally confound observational studies, provider-level studies are also complicated by substantial noise. Many patient and other factors influence most treatments and outcomes of interest, and most providers treat only a limited number of patients in a particular time

period. In addition, many dimensions of quality – numerous different processes of care and outcomes – may be relevant to the judgments of patients and their physicians about quality of care. As a result, virtually all estimates of provider quality reported to date are very imprecise and incomplete, and many experts have expressed doubt whether meaningful empirical quality assessment is feasible at all for most aspects of medical care (see McClellan and Staiger, 2000, for a more detailed discussion). In this paper, we not only develop and apply methods for obtaining more precise estimates of many dimensions of treatment at the hospital level simultaneously; we also develop and apply methods for summarizing these many dimensions accurately and determining whether these differences in practice patterns help explain the substantial differences in health outcomes that we also observe across providers. Thus, our methods help address many of the difficult issues in evaluating health care providers.

We apply our GMM method to evaluate hospitals and the effects of the treatments they provide for elderly Americans with heart attacks. We use detailed medical chart review data and linked Medicare administrative records from the Cooperative Cardiovascular Project on over 170,000 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. Second, we find that the GMM method, which uses this variation in treatment intensity across hospitals to estimate treatment effects, yields treatment effect estimates that are quite similar to recent IV estimates (e.g., McClellan et al., 1994) but that are considerably more precise. In addition, the GMM estimates appear to

be robust to controlling for detailed information on patient severity. Finally and most importantly, we are able to reach a number of conclusions about how medical practices for heart attack care differ across hospitals, and the consequences of these differences for heart attack mortality and for providing insights into why some hospitals are able to achieve much better mortality outcomes than others.

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 relevant treatment effects precisely. Patients and their physicians may believe, based on limited evidence, that a particular therapeutic alternative may be more beneficial for them; especially if the treatment decisions have potentially important implications for survival (e.g., major treatments for heart attacks), they may rationally believe that randomization is not worthwhile.

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

¹Trials 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).

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 undergoing 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 smaller mortality effects in patients undergoing procedures, and the effects appeared to be due at least in part to 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, 2000). 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 somewhat 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 heart attack is more controversial. Trials in the early 1990s generally found no difference in outcomes; however, some recent trials have concluded that primary angioplasty leads to lower mortality, and many cardiologists have expressed concerns about the safety of thrombolytics in elderly patients who are more prone to bleeding complications (the randomized trials of thrombolytic drugs have mostly been performed in nonelderly patients, without many comorbid illnesses). Beta blockers, aspirin, and (for the subset of 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; at least in the 1994-95 period, no randomized trials had found a clear benefit to using calcium channel blockers shortly after heart attack. 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. Because most of these trials were limited to nonelderly patients with few comorbid diseases, however, their relevance to elderly AMI patients has been questioned.

C. Evidence on Hospital Quality Differences

Evidence on the quality or effectiveness of particular hospitals in caring for AMI patients has been even more difficult to obtain. Because hospitals treat only a limited number of AMI patients and because many patient and environmental factors influence AMI outcomes, standard comparisons of mortality rates across hospitals even after adjusting for observable patient differences are imprecise (Thomas and Hofer, 1999). They are also very likely to identify hospitals as high-mortality “outliers” by chance, rather than as a result of true or systematic differences in quality. Our previous work (McClellan and Staiger, 1999a, 1999b, 2000) developed and applied methods to distinguish true or systematic variation in treatments and outcomes

across hospitals from such random variation. Our methods are closely related to Bayesian methods for “smoothing” hospital estimates, but they make use of the fact that many dimensions of care are related to each other. For example, hospitals that have lower mortality rates in one year tend to have lower mortality rates in other years, and hospitals that perform better in some dimensions of care tend to perform better (and occasionally worse) in other dimensions. We showed that such correlations can be used to reduce the noise problem substantially in estimating hospital effects.

However, our previous work did not provide any evidence on how lower-mortality hospitals achieved better outcomes; nor did it provide any evidence on whether the observational data underlying our estimates of hospital quality could provide further insights into the effects of particular medical treatments. These are important issues. There is considerable evidence that hospitals often fail to provide care that has been shown to be effective (e.g., Asch et. al., 2000). Yet, because of problems in identifying differences in hospital practices with any useful degree of precision, such studies are often reported only at the level of broad groups of hospitals (e.g., low- and high-volume hospitals, or teaching and nonteaching hospitals) or at the level of geographic regions, entire states, or the country as a whole (e.g, Allison et. al, 2000). At such broad levels of aggregation, it is difficult to develop policy or management interventions that encourage appropriate hospital practices and discourage ineffective ones. The inability to identify important differences in practices and their consequences at the hospital level thus inhibits the implementation of programs to improve quality of care.

3. Methods

A. Overview and identification

A 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 affects 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 at the time of admission. 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, this approach has two important drawbacks. One 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.

A second important drawback of this approach is that some important aspects of treatment are hard to measure, but are likely to be correlated with a hospital's treatment intensity for measured treatments, leading to bias. For example, hospitals that tend to treat heart attack patients with aggressive surgical procedures (bypass, angioplasty) may also have a more skilled nursing staff, better technology available in their Intensive Care Units, or provide additional follow-up care after the patient is discharged. In this case, the estimated treatment effect may be the result of related treatments that were not included in the model. Therefore, we develop a factor model that summarizes the variation in treatment intensity with a smaller-dimensional set of factors, and then relate these factors to patient outcomes. This approach acknowledges that one cannot estimate the effect of individual treatments, but retains the ability to relate observed differences in practice patterns to patient outcomes.

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 effect models (or factor models) 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 (or factor) 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 $1 \times K$ vector of treatments (T_{ij}) and a $1 \times L$ vector of patient characteristics measured at admission (X_{ij}) according to the following equation:

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

The parameters of interest are β (a $K \times 1$ vector of treatment effects) and Π (a $L \times 1$ vector of covariate effects). We decompose the error into two parts: A hospital component (α), which as discussed above is assumed to be uncorrelated with T ; and a patient component (ε) that may be correlated with T . Patient characteristics at the time of admission (X) are assumed to be uncorrelated with ε .

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:

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

Where equation 2 is a multivariate regression where Γ is a $L \times K$ matrix of parameters, and δ_j is a $1 \times K$ 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 v_{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:

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

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

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 v 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 (1999a,b, 2000) 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.

C.1. 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 $M_j = \langle \hat{\mathbf{q}}_j, \hat{\mathbf{d}}_j \rangle$ be the $1 \times (K+1)$ vector of estimated fixed-effect parameters for hospital j , and let S_j be the OLS estimate of the $(K+1) \times (K+1)$ 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 and S_j are independent across hospitals.

C.2. 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:

$$(4) \quad M_j = \langle \hat{\mathbf{q}}_j, \hat{\mathbf{d}}_j \rangle = \langle \mathbf{q}_j, \mathbf{d}_j \rangle + \mathbf{z}_j$$

Where $\langle \mathbf{q}_j, \mathbf{d}_j \rangle = \langle \mathbf{d}_j \mathbf{b} + \mathbf{a}_j, \mathbf{d}_j \rangle$, and \mathbf{z}_j is the $1 \times (K+1)$ vector of estimation error which is mean zero and uncorrelated with $\langle \mathbf{q}_j, \mathbf{d}_j \rangle$. This implies that:

$$(5) \quad E(M_j' M_j) = E \left(\langle \mathbf{q}_j, \mathbf{d}_j \rangle' \langle \mathbf{q}_j, \mathbf{d}_j \rangle \right) + E(\mathbf{z}_j' \mathbf{z}_j)$$

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, $E(\mathbf{z}'_j \mathbf{z}_j)$. Therefore, we have:

$$(6) \quad E(M'_j M_j - S_j) = E\left(\langle \mathbf{q}_j, \mathbf{d}_j \rangle' \langle \mathbf{q}_j, \mathbf{d}_j \rangle\right)$$

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:

$$(7) \quad E(M'_j M_j - S_j) = \begin{bmatrix} \mathbf{b}' \Omega_d \mathbf{b} + \mathbf{s}_a^2 & \mathbf{b}' \Omega_d \\ \Omega_d \mathbf{b} & \Omega_d \end{bmatrix}$$

where $\Omega_d = Var(\mathbf{d}_j)$, and $\mathbf{s}_a^2 = Var(\mathbf{a}_j)$.

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 (\mathbf{s}_a^2). In particular, we estimate the parameters by setting the theoretical moments (the right hand side of equation 7) equal to a 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:

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

where α_j is assumed to be independent of δ_j . Thus, if we observed θ_j and δ_j without any estimation error, 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{q}_j and \hat{d}_j . Thus, rather than estimating β with the usual least squares formula (ignoring the weights) of $(\hat{d}'\hat{d})^{-1}(\hat{d}'\hat{q})$, we use estimates of the moment matrices that correct for the measurement error. For example, $\frac{1}{N}(\hat{d}'\hat{d})$ overstates the variance in treatment across hospitals ($E(\mathbf{d}'\mathbf{d}) \equiv \Omega_d$) because of the estimation error in \hat{d} . Therefore, our method “subtracts off” the estimation error and estimates the variance in treatment across hospitals (Ω_d) with $\frac{1}{N}(\hat{d}'\hat{d}) - \bar{S}_d$, where \bar{S}_d is the average estimation error variance for the treatment fixed effects (\hat{d}_j).

More generally, our GMM method provides a simple and intuitive framework for addressing the problem of weak instruments. In our application, the problem of weak instruments corresponds to the problem that θ_j and δ_j are observed with substantial (and correlated) estimation error, whereas the conventional asymptotic properties of 2SLS rely on this estimation error disappearing asymptotically. The presence of estimation error in θ_j and δ_j biases the 2SLS estimate of β (toward OLS, because the correlation in the estimation error is in that direction) and biases standard error estimates of 2SLS (usually downward, because the estimation error overstates the variance of δ_j). Our GMM method relies on the hierarchical structure of the data (with multiple observations for each hospital) to generate estimates of the variance and covariance in the estimation error, and then uses these estimates to correct for the biases of

2SLS. In a similar approach, Chamberlain and Imbens (1996) also exploit this hierarchical structure to derive alternative IV estimators that are robust to weak instruments. Other approaches to overcoming weak instruments are also easy to interpret in this framework. For example, Angrist and Krueger (1995, 1999) take a split-sample approach, estimating θ_j and δ_j from different samples. This results in independent estimation error and, therefore, a bias toward zero (rather than OLS). Standard errors still have to be adjusted in their approach, since the estimation error overstates the variance of δ_j .

D. Factor Models

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.

Given the difficulty in measuring all relevant treatments in most medical situations, one would like a model that is able to estimate the effect of observed differences in practice patterns on patient outcomes, yet at the same time acknowledges that one cannot attribute the effect to individual treatments. The approach that we adopt is to use a factor model in which all treatments (both observed and unobserved) depend on a smaller-dimensional set of factors, and then these factors (rather than the treatments themselves) affect patient outcomes. Thus, we assume that mortality rates at a hospital depend on overall practice patterns (summarized by the factors) that can be identified from variation in the intensity of observed treatments.

More specifically, suppose that we replace equation 8 with a factor model in which both a hospital's treatment intensity (δ_j) and patient outcomes (θ_j) depend on a small-dimensional set of latent factors:

$$(8') \quad (\theta_j, \delta_j) = F_j \Lambda + \nu_j$$

where F_j is a $1 \times J$ vector of factors with dimension less than the number of treatments ($J < K$), Λ is a matrix of coefficients (the factor loadings) giving the effect of the factors on the treatments and outcomes, and ν_j is a vector of independent residuals with $\text{Var}(\nu) = \Psi$ a diagonal matrix. Without loss of generality we normalize the factors so that $\text{Var}(F) = I$. Thus, the factor loadings (Λ) can be interpreted as the increase in treatment intensity or patient outcomes associated with a one standard deviation increase in each factor.

The factor structure (equation 8') implies that $E(M_j' M_j - S_j) = \text{Var}(\theta_j, \delta_j) = \Lambda' \Lambda + \Psi$. Thus, the parameters of the factor model (Λ, Ψ) can be estimated by Optimal Minimum Distance (OMD) methods (Chamberlain, 1984), i.e. by choosing the parameters so that the theoretical moment matrix, $\Lambda' \Lambda + \Psi$, is as close as possible to the corresponding sample moments from the (weighted) sample average of $M_j' M_j - S_j$. More specifically, let d_j be a vector of the non-redundant (lower triangular) elements of $M_j' M_j - S_j$, and let $\Delta = g(\Lambda, \Psi)$ be a vector of the corresponding moments from the true moment matrix. Then the OMD estimates of (Λ, Ψ) minimize the following objective function:

$$(9) \quad q = N \left[\bar{d} - g(\Lambda, \Psi) \right]' V^{-1} \left[\bar{d} - g(\Lambda, \Psi) \right]$$

where V is the sample estimate of the covariance matrix for d , and \bar{d} is the sample average of d , and both estimates are weighted by the number of admissions at each hospital.

If the factor model is correct, the value of the objective function, q , will be distributed $\chi^2(p)$ where p is the degree of over-identification (the difference between the number of elements in d and the number of parameters being estimated). Thus, q provides a goodness of fit statistic that indicates how well the factor

model fits the actual covariance in the data. However, there is considerable evidence that this goodness of fit statistic tends to over-reject in practice (Hansen et. al, 1996; Stock and Wright, 1996). Thus, to avoid over-fitting the data we will apply this statistic conservatively.

There is similar evidence suggesting that conventional standard errors for OMD estimates tend to be too small, and that confidence intervals derived directly from the objective function tend to have better coverage (Hansen et. al, 1996; Stock and Wright, 1996). Therefore, we derive standard errors from a delta-method estimate for the variance of the parameters, where the delta-method estimates are derived directly from the objective function. In particular, note that \bar{d} is simply a vector of sample averages with $\text{var}(\bar{d})=V/N$. Thus the OMD parameter estimates are an implicit function of the estimated parameter vector (\bar{d}), as defined by the objective function (equation 9). Letting $\theta=(\Lambda,\Psi)$, by the delta method we have:

$$\text{Var}(\hat{q}) = \frac{\partial \hat{q}}{\partial \bar{d}} \left(\frac{V}{N} \right) \frac{\partial \hat{q}'}{\partial \bar{d}'},$$

where we estimate the derivatives of the OMD estimates with respect to \bar{d} numerically. Standard errors based on this approach were generally 2-10 times larger than the conventional OMD standard errors.

There are a number of advantages of the factor model. Most importantly, if some treatments are unobserved, the remaining parameters (i.e. the components of Λ and Ψ that correspond to the observed treatments and outcomes) can still be consistently estimated. Thus, this approach avoids the omitted variable bias inherent in estimating treatment effects by assuming that it is the variation in the latent factors, rather than specific treatments, that influences outcomes. In addition, the assumptions of the factor model can be tested through the goodness of fit statistic. This goodness of fit statistic is a global test of the over-identifying assumptions in the factor model, which imply that consistent parameter estimates could be obtained using any just-identified subset of the observed treatments. Finally, the factor model summarizes the multi-dimensional

variation in treatment patterns across hospitals with only a few factors, and thereby focuses attention on a few simple dimensions on which practice patterns vary.

There are also a number of important limitations of the factor model. First, the estimated factor loadings (7) are only unique up to an ortho-normal rotation. We report estimates from the varimax rotation, which is commonly used in factor analysis because it maximizes the amount that each outcome is associated with a unique factor, thus helping with interpretation of the factors (Johnson and Wichern, 1998). A second limitation of the factor model is that the appropriate number of factors is not obvious. We choose the number of factors empirically, based on a combination of statistical significance and robustness of results with the incremental addition or removal of factors. A final limitation is that, even though we have sought to isolate the consequences of incremental variations in the use of additional treatments to the extent our empirical evidence allows, it remains possible (and perhaps likely) that each of the observed “practice patterns” described by the factors are related to other, unobserved treatments. Nonetheless, the factor model provides a more quantitative proxy for overall practice patterns that, like the complex characteristics of patients that influence treatment, can only be partially observed in the data.

4. Data

To illustrate the application of our GMM methods to evaluating treatments and understanding the causes of differences in outcomes among medical care providers, 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 185,000 elderly 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 provide 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. Our analysis uses the following variables:

Treatment measures. CCP data permit construction of a broad range of medical treatment measures. Diagnostic treatments included whether or not a set of laboratory tests were performed (e.g., blood counts and chemistries) and the outcome of these tests; whether or not a chest X-ray, electrocardiogram (EKG), and other relatively routine diagnostic tests were performed, and their outcomes; as well as the performance and findings of some more specialized diagnostic tests including cardiac catheterization and echocardiography during the hospital admission. A range of treatments with commonly-used drugs after AMI were also reported, including aspirin, beta blockers, ACE inhibitors, calcium channel

blockers, heparin, IV nitroglycerin, and lidocaine. As we describe in more detail below, some of these treatments (but not all) have been shown in clinical trials to reduce heart attack mortality. CCP data also report whether or not the patient received a thrombolytic or “clot-busting” drug, which is intended to head off a heart attack in progress by dissolving the blood clot and associated blockage that caused the heart attack. While thrombolytics have been shown in numerous clinical trials to reduce heart attack mortality significantly, serious hemorrhage (including GI and intracranial bleeding) is a complication of treatment, particularly in the elderly.

From the linked administrative claim data, we determined whether patients received intensive cardiac procedures within 90 days after AMI: cardiac catheterization, PTCA (angioplasty), and cardiac bypass (CABG) surgery. We also constructed a measure of the use of “primary” angioplasty within 1 day of AMI. This intensive treatment seeks to re-open the blockage causing a heart attack within hours of its occurrence, in order to restore blood flow to the affected heart muscle before it dies, thereby limiting the damage from a heart attack. Thus, it is viewed clinically as an alternative to the use of thrombolytic drugs. While it avoids the bleeding complications associated with thrombolytics, it is much more resource-intensive (e.g., “on-call” specially-trained cardiologists and technicians are required) and may be associated with some complications of its own (e.g., re-blockage).

Outcome measures. We use death dates from all causes validated by the Social Security administration to calculate survival times from the initial hospitalization. Based on these dates, we compute 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 covariates:

- (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, heart rate, mean arterial pressure, respiratory rate, and whether the patient is verbally oriented.

For detail on how these variables are constructed, see McClellan and Noguchi (2000); 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 adjusted for the detailed list of patient covariates just described (i.e., the estimates of hospital fixed effects from equations 2 and 3). All statistics have been weighted by the number of new AMI admissions at each hospital. Mortality for heart attack patients is relatively high, with 17% in the first 30 days, and 30% in the first year following the heart attack. Estimated mortality rates also vary considerably across hospitals, e.g., the standard deviation of 30-day mortality across hospitals is 6 percentage points. Because the average hospital admitted only 43 elderly heart attack patients during the study period, a substantial part of this variation is likely due to estimation error in each hospital's estimate. Around 47% of patients receive catheterization within 90 days after the heart attack, with over 18% of these patients going on to have primary (one-day) angioplasty (8.7% of the total) and an additional 23% going on to have angioplasty within 90 days (11.0% of the total). Around 1/3 subsequently undergo bypass surgery (16.0%). Treatment rates for the medications we study, beta blockers (46.7% during hospitalization (and a virtually identical rate at discharge) and aspirin (78.5%), are also substantial – 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 7.1 (bypass) to around 10-15 percentage points (catheterization, aspirin, beta blockers).

Following McClellan et al. (1994), we construct “differential distance” variables for each patient, based on the difference between the distance to the nearest hospital able to perform catheterization, less the distance to the nearest hospital not performing catheterization. We construct similar variables for high volume hospitals (75+ patients) and hospitals able to perform primary angioplasty. Distance estimates are measured between population centroids of the zip code of residence and the zip code of each hospital. This variable has been shown to perform well as an instrument for catheterization, in that it is correlated with receiving catheterization (patients nearer a catheterization hospital are significantly more likely to receive this treatment), and it does not appear to be related substantially to differences in patient health characteristics that directly influence outcomes. For a more details on the construction, use, and validation of this instrumental variable, see McClellan et al. (1994) and McClellan and Noguchi (2000). Making use of our GMM signal extraction methods, we also construct IV estimates based on zip-code differential distance groups defined by dividing the three distance variables into 0.5-mile increments in urban areas, and 4.0-mile increments in rural areas.

5. Results

In this section we report our GMM estimates of differences in outcomes and medical practices across hospitals, and of the relationship between hospital- and zip-level differences in medical practice patterns and mortality outcomes. We begin by reporting our GMM estimates of the variance in outcomes and treatment intensity across hospitals. These estimates are of independent interest, since they document the extent of practice and outcome variations across hospitals after accounting for measured differences in

patient mix as well as apparent differences due to statistical noise. We then report alternative 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, comparing our GMM results to the effect of catheterization estimated using simple OLS methods and previously-reported differential-distance IV methods (McClellan et al., 1994; McClellan and Noguchi, 2000). After discussing the limitations of all of these models of specific treatment effects in comparison to methods that account for hospital- or zip-level noise and the correlations among many medical treatments, we present our GMM-IV estimates of the effects of alternative practice patterns, using individual hospitals and zip code differential distance groups as instruments.

Estimates of Variance and Correlation in Treatment and Outcomes Across Hospitals

We find substantial variation in both treatment rates and mortality rates across hospitals, after adjusting for observed differences in case mix and random hospital-level variation. Tables 2 and 3 report the GMM estimates for the variation across hospitals in measures of mortality and treatment intensity, respectively. Along the diagonal of each table, we report estimates of the standard deviation in each measure of hospital quality, while the off-diagonal elements of each table report the estimated correlations between pairs of 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: 30-day mortality and 1-year mortality. The estimated standard deviation in mortality rates across hospitals is substantial at 3.2 percentage points for 30-day mortality (relative to a base of 17.2 percentage points). 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. Between 30 days and 1 year, very little additional systematic variation mortality across hospitals is observed: the standard deviation for 1-year mortality is also 3.2 percentage points (relative to a base of 29.5%). Thus, additional mortality that occurs after 30 days is essentially “random noise” from the standpoint of the quality of the initial treating hospital. The variation in mortality rates across hospitals is also highly correlated across the mortality measures, with a correlation of over 0.9 between the 30-day and 1-year measures. Thus, differences in mortality across hospitals that appear in short-term mortality measures persist through 1-year mortality. These results are consistent with our earlier findings (McClellan and Staiger, 2000), which concluded from a less clinically-detailed time series that the systematic variation in mortality across hospitals treating AMI patients approached its maximum within the first month of treatment; observation of longer-term mortality outcomes added little to the quality evaluation.

As we noted in that study, this result is also consistent with a range of clinical trial results showing substantial and persistent benefits of many treatments acutely after AMI. However, we did not have data with enough clinical details to allow us to observe what those particular treatments were. Here, our chart-review data allow us to evaluate the extent to which a broad range of treatments differed across hospitals.

Table 3 reports similar GMM estimates of the true variation across hospitals for nine selected treatment measures. Again, the standard deviation in treatment rates across hospitals is large for most of the measures, e.g., 13 percentage points for catheterization within 90 days, 9 percentage points for primary angioplasty (1-day PTCA), 9 percentage points for aspirin, 13 percentage points for beta blockers, and over 7 percentage points for thrombolytic use. Thus, treatments differ substantially across hospitals, even for inexpensive drugs. In most cases, treatment rates are positively correlated, suggesting that many dimensions

of hospital practice are complements for each other. Hospitals that use one surgical treatment aggressively tend to use other surgical treatments aggressively as well. Hospitals that are more likely to perform catheterization are also much more likely to perform bypass surgery and angioplasty (correlations of 0.6 and 0.7 respectively). Hospitals that are more likely to perform bypass surgery are also more likely to perform angioplasty; thus, these alternative revascularization procedures are not substitutes in practice. The strong correlations suggest that it may be difficult to distinguish the effects of catheterization (if any) from bypass and angioplasty in our estimates based on variation across hospitals.

In addition, hospitals that use intensive procedures more often also appear more likely to use some potentially beneficial drug therapies like aspirin (correlation with catheterization use of over 0.3). Thus, some of the drug treatments that have been shown to reduce heart attack mortality in clinical trials are in general more likely to be used at hospitals that treat patients more intensively in other respects, again complicating efforts to distinguish the effects of particular medical practices. One exception is thrombolytic drug use, which is negatively correlated with the use of the catheterization and quite negatively correlated with the use of primary angioplasty. Thus, thrombolytic drugs are to some extent substitutes for the use of cardiac procedures to restore blood flow to the heart soon after a heart attack. Another exception (not shown in the table) is counseling for smoking cessation, which differs from the other treatments included in the table in that it is not a specific “medical” therapy but rather a psychosocial activity. This treatment appears less likely to be used (or at least reported) at the hospitals that provide the most intensive treatment in other respects.

Estimates of Treatment Effects for Catheterization

Table 4 reports estimates of the treatment effect (β) for simple versions 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. As suggested by the strong correlation between catheterization and other invasive revascularization procedures in Table 3, these results might be viewed as alternative estimates of the effects of generally more “intensive” heart attack treatment. Each column reports estimates for two mortality windows: 30-day and 1-year. The left two columns report estimates that control for the full set of patient covariates available in the CCP data. The right two columns report estimates of mortality effects for the same time intervals that only control for a limited set of demographic variables (gender, race, age, urban residence) commonly available in claims data. The first row reports our GMM estimates based on true hospital-level variations in catheterization use and mortality, the second row reports IV estimates based on differential distance to a hospital performing catheterization (see, e.g., McClellan et al., 1994), and the third row reports OLS estimates that do not account for unmeasured differences in patient selection for catheterization.

If the assumptions required for consistency of the GMM estimator are correct, then the GMM results should be consistent but more efficient than the IV estimates, because the IV estimates use less of the systematic variation in treatment rates across hospitals. In contrast, the OLS estimates will most likely be biased toward overstating the treatment effect due to selection (patients who die soon after admission and patients who have other, potentially hard-to-measure comorbidity or severity characteristics are less likely to receive catheterization). This is the pattern observed in the estimates that control for the full set of patient covariates (the left half of the table). The GMM point estimates are not significantly different from the IV estimates, but they are considerably more precise, with standard errors 3-4 times smaller than the IV 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 report estimates from a similar set of models to the left panel, but with controls for a much more limited set of patient covariates (demographics only). 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, and also provide some suggestive evidence on the extent of residual biases in estimates based on the alternative methods. The results demonstrate that, at least in this application to heart attack mortality, the GMM and IV methods are relatively 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. Although the differences are not significant, the GMM estimates show a slightly different time pattern compared to the IV estimates, with the effect magnitudes growing larger over time at a more rapid rate, especially at one year. In our discussion of Table 6 below, we return to the question of whether the hospital-based GMM approach may contain a relatively small amount of residual bias. In contrast to both of these IV approaches, OLS estimates are extremely 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. Thus, it seems likely that the OLS estimates (even with detailed covariates included) that have been widely reported in the medical literature, are much more likely to be biased as a result of residual unmeasured differences in patient selection.

Overall, the results in Table 4 support three 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). Thus, provided that patient sorting to hospitals based on unobserved severity is not a significant problem, our method substantially expands the IVs available for estimating treatment effects with observational data. Second, as previous IV estimates have suggested, OLS estimates appear to be biased significantly toward overstating the effects of intensive cardiac treatments, even in analyses that control for clinically detailed patient covariates. Thus, at least for intensive procedures after heart attack, OLS and similar methods based on controlling for observed differences between treated and nontreated patients appears to be a fundamentally flawed approach to 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, GMM methods appear to be a practical approach for estimating treatment effects using commonly-available claims datasets (provided that treatments are measured in the claims data).

However, even though the GMM and IV methods appear to perform much better in resolving the problem of biased estimation due to patient selection for treatment, interpreting estimates based on all of these methods is limited by the fact that more use of intensive procedures like catheterization is in general significantly correlated with a broad range of other potentially beneficial treatments. In other words, while the IV and GMM estimates of catheterization effects provide some insights about the overall effect of more “intensive” medical practice on heart attack mortality, the estimated effect is unlikely to be the result of catheterization per se, but rather of other treatments that are correlated with catheterization. Indeed, McClellan et al. (1994) and McClellan and Noguchi (2000) found that most if not all of the apparent effect of catheterization in IV analysis could be explained by the correlation of catheterization use with treatments other treatments. We now turn to estimating models using our GMM methods that account for a much more comprehensive set of heart attack treatments.

Estimates of Treatment Effects in More General Models

In this section we estimate factor models of mortality, which include surgical treatments (bypass, angioplasty) and drug treatments (aspirin, beta blockers) in addition to catheterization. A very important advantage of our method is that it more easily extends to the analysis of many treatments. In contrast, while IV estimates as in McClellan et al. (1994) have proven feasible for analyzing several treatments, they are difficult to apply to models with many treatments, because it is difficult to identify additional strong instruments that generate significant independent variation in the additional treatments. In Table 5, we make use of the large variation across hospitals in a broad range of heart attack treatments, including but not limited to the illustrative treatments described in Table 3. In particular, we use factor models (equation 8) to estimate the mortality effects of underlying factors that summarize the impact on mortality of groups of treatments that tend to vary together.

Tables 5 and 6 report mortality effect estimates from GMM factor analyses that evaluate the relationship between the mortality outcomes and “practice patterns” estimated from the hospital-level or area-level variations in medical practices. Table 5A reports results of factor models for 30-day mortality, and table 5B for 1-year mortality, using hospital IVs. Table 6 reports results of factor models for 30-day mortality using the zip-code IVs. All models were estimated with a full set of demographic, comorbidity, and disease severity controls included (i.e., analogous to the left half of Table 4). In each table the estimates of the association of each factor with the mortality outcome and with the use of each treatment is listed. The estimated factor effects on outcomes and the treatment loadings are reported in percentage points (with standard errors in parentheses); these estimates show the effect of a one-unit change in the factor, which is normalized to have a standard deviation of 1. We also report the proportion of variation in mortality and

each of the treatments that is explained by the factors, and the proportion of the total variation in mortality and all treatments explained by each factors. The factors have been ordered from left to right in the table according to the proportion of total variation each factor explains. This is a somewhat arbitrary ranking, as it depends on the particular treatments included in our empirical analysis. We comment further on this issue below.

The factor loadings reported in Tables 5A and 5B indicate that each factor was associated with significantly greater rates of performance of some aspects of heart attack treatment, and each was associated with at least marginally lower mortality(the first row of the Table labeled “Mortality Effect”). Moreover, the factor loadings were quite similar in the 30-day and 1-year models. Factor 2 was associated with the largest mortality reduction. The treatments loading most strongly on this factor were aspirin use (6.6 percentage-point loading for 30-day mortality, and similar loading for 1-year mortality), beta blocker use (9.0 percentage-point loading), and ACE inhibitor use (3.6 percentage-point loading). The use of each of these drugs has been shown in highly-regarded clinical trials in the 1980s and early 1990s to be associated with significantly lower mortality after heart attack. Thus, this factor might be regarded as an “evidence-based best practice” factor – hospitals scoring highly on it generally appear to be more effective in implementing the results of clinical trials in heart attack care. This factor explains by far the largest share of variation in 30-day and 1-year mortality across hospitals. However, it accounts for only around 20% of the total systematic variation in between-hospital mortality. Thus, other processes or aspects of care not related to “evidence-based” best practice account for most of the variation in outcomes across hospitals.

Factors 1, 4, 5, and 6 are all associated with more aggressive use of intensive cardiac procedures, beginning with cardiac catheterization. Factor 5, which has the largest mortality effect of the intensive-procedure practice patterns, is associated with more aggressive use of catheterization and primary

angioplasty, as well as much less use of thrombolytic drugs. As noted above, primary angioplasty is a more intensive alternative to the use of thrombolytic drugs in restoring blood flow immediately after a heart attack. Thus, this factor captures a more aggressive style of cardiologist intervention, and is associated with considerably lower mortality. A one standard-deviation higher score on this factor is associated with a 2.4 percentage-point higher rate of use of PTCA, a 4 percentage-point lower rate of use of thrombolytic drugs, and 1 percentage-point lower mortality rate. This large mortality effect suggests that primary angioplasty is a preferable treatment to thrombolytic drug use, and that thrombolytic drugs may even be harmful or at least should be used judiciously. However, because the magnitude of the effect of this factor is too large to be clinically plausible, it is likely that the associated practices are in part proxying for hospitals that are providing higher-quality care in other unmeasured dimensions. For comparison, Factor 1 is associated with much more primary angioplasty use (loading of 7.1 percentage points) and with somewhat less thrombolytic use (loading of -2 percentage points), but this factor has only a modest effect on 30-day and 1-year mortality. Together, these factors indicate that hospitals using more primary angioplasty instead of thrombolysis tend to have better outcomes, but that it is difficult to estimate a precise effect of primary angioplasty because of its apparent correlation even in our detailed analysis with other unmeasured practices. Some of the hospitals that perform more primary angioplasty and that are selective in their use of thrombolytic drugs tend to have particularly good outcomes.

Factor 4, which involves more use of catheterization and bypass surgery (and to some extent primary angioplasty), is associated with an intermediate mortality reduction at 30 days, around 0.7 percentage points. However, the magnitude of the effect of this practice patterns increases over time, to 1 percentage point at 1 year, when the longer-term benefits of major surgery might be expected to be observed. Factor 6 is associated with much greater use of catheterization as well as greater use of non-“primary” angioplasty. It is

also associated with somewhat greater use of thrombolytics and less use of primary angioplasty. Thus, this practice pattern describes hospitals that tend to use thrombolytic drugs to restore blood flow initially, but that provide more intensive followup treatment. This factor is associated with a modest beneficial mortality effect at 30 days and 1 year, suggesting that thrombolytic drug use is an effective alternative to primary angioplasty, provided that followup angioplasty is performed aggressively.

Factor 3 is associated with somewhat greater use of primary angioplasty (3.3 percentage-point loading) and a greater likelihood of performing and reporting a range of diagnostic tests on heart attack patients, including various blood tests, a chest X-ray, and an electrocardiogram. While these diagnostic tests are unlikely to lead to better outcomes in themselves, their greater performance may be an indicator of good “basics” of medical care, including perhaps better documentation practices. These good “basics” are also associated with a minimal (and insignificant) beneficial mortality effect.

We have estimated a range of other models that included additional treatments, but these models provide few additional insights into our practice pattern analysis. For example, use of specific drugs at discharge from the hospital (i.e., use of aspirin and beta blockers at discharge as indicators of the intensity or quality of post-hospital care) load very similarly to use of the same drugs during hospitalization. Use of other “basic” diagnostic tests loaded similarly to the diagnostic tests in Factor 3. Use of smoking cessation counseling and some other medical treatments including IV nitroglycerin, lidocaine, and calcium channel blocker use was not strongly associated with other treatments nor with 30-day or 1-year mortality. As we noted above, these drug therapies would not be expected to have substantial mortality effects based on existing clinical evidence. While smoking cessation would be expected to reduce long-term mortality, it is possible that effects are not apparent by one year, or that counseling is heterogeneous enough that it is

difficult to discern an effect. Finally, we were not able to identify any other treatments that were strongly associated with the notable mortality effect in Factor 5.

Table 6 presents analogous factor analysis results for our zip-based GMM IV models, that is, models using variation in treatments and outcomes across zip code differential distance groups rather than hospitals. This analysis allows us to explore whether relatively minor but still potentially important biases in patient selection at the hospital level may complicate the interpretation of the estimates in Table 5 (see the discussion of the GMM and IV results of Table 4, above). Because of the substantially lower signal (systematic) variation associated with the zip code groups – the magnitude of the signal variation in zip mortality is only around one-third the size of the magnitude of the signal variation in hospital mortality – we were only able to identify two practice patterns that explained a significant proportion of the variation in treatments and outcomes across zips.

Despite this limitation, the estimates of practice patterns and their effects bear some important differences and similarities to the results presented using hospital-level analyses. One similarity is that the most important practice pattern for explaining mortality differences, Factor 1, again consists of “evidence-based best practice,” including more use of ACE inhibitors, aspirin, and especially beta blockers. This practice pattern explains a relatively large share (over one-third) of the between-zip variations in outcomes. A second similarity is that Factor 2 shows that more aggressive use of intensive procedures, particularly catheterization and primary angioplasty, is associated with lower mortality. However, the magnitude of this “intensive procedure” effect is smaller than the effects associated with similar factors in Table 5, and is no longer statistically significant. These results suggest that, even though the hospital IV estimates of Table 5 appear to be relatively free of selection bias, the hospitals with the most aggressive use of intensive procedures may be attracting heart attack patients with somewhat better prognoses in dimensions that we

are not able to measure, even with very detailed medical chart reviews. Thus, the benefits of intensive practices suggested by some of the factors in Table 5 may be somewhat overstated.

Table 7 reports goodness-of-fit tests for the models reported in Tables 5 and 6. As discussed previously, such tests tend to over-reject models, suggesting that these results should be interpreted conservatively. The Table shows that the addition of a seventh factor in the hospital IV models (Table 5) led to a statistically significant but relatively modest improvement in goodness of fit. Moreover, the addition of the seventh factor had minimal effect in terms of predictive power. As the final row of Tables 5A and 5B shows, the sixth factor explained less than 5% of the total variance in hospital-level treatments and outcomes. Models with additional factors explained less than an additional 2% of the variation, and so did not alter any of our substantive conclusions. For our zip IV analysis, the addition of a third factor had a similar, minimal effect on explanatory power.

Overall, our results indicate that our GMM estimation methods appear to provide a much clearer and more precise way of interpreting the evidence provided by observational data on health care providers, treatments, and health outcomes. We are able to: remove the substantial noise that has confounded analyses at the level of individual providers or small groups; summarize concisely how a large number of medical practices differ across hospitals; and estimate relatively precise treatment effects in models that account for many covariates and the joint use of many treatments. Moreover, our strongest practice pattern effects are tightly consistent with biomedical knowledge. That is, greater use of aspirin, beta blockers, and ACE inhibitors – the treatments found to have the most notable benefits for heart attack outcomes in the decade preceding our analysis – is clearly associated with lower mortality. However, our results also show that better adherence to “evidence-based” guidelines alone does not account for the bulk of the large variations in outcomes across hospitals or geographic areas. Thus, the evidence we develop on the benefits of a range

of other medical practices, particularly those involving intensive cardiac procedures, may provide important supplements to evidence from clinical trials.

5. Conclusion

Estimating the effects of medical and surgical treatments in a valid manner is the foundation of evidence-based medicine. Despite the fact that the number of clinical trials performed is rising rapidly, most medical practices – and particularly intensive practices that vary considerably across hospitals and areas – have not been evaluated in formal clinical trials. Moreover, clinical trials will never be able to provide valid evidence on differences in performance across health care providers. In this paper, we have developed and applied a set of methods for identifying the relationship between differences in medical practices and health outcomes using observational data, while avoiding the usual problems of selection bias in observational analysis. Our methods also strengthen the methods we have developed and applied in previous work (e.g., McClellan et al., 1994; McClellan and Staiger, 2000) by allowing far more precise estimation of both treatment effects and the specific ways in which medical practices differ across hospitals.

On a substantive level, our analysis of hospital practices shows that the effects of drug treatments (aspirin, beta blockers, ACE inhibitors) and surgical treatments (bypass, angioplasty) on survival are both significant and substantial. We also find some suggestive evidence that more aggressive use of intensive cardiac procedures, better performance and documentation of basic diagnostic procedures, and (in conjunction with certain other procedures) more use of thrombolytic drugs are also associated with somewhat better outcomes. However, because of the more limited zip-based variation in practices, we have not been able to confirm fully that the results on intensive procedures are robust to additional steps to control for residual selection biases.

Substantial variation continues to exist across hospitals in the rates at which heart attack patients are provided with all of these treatments. Our results provide considerable insights into which practices tend to go together, thus improving our understanding of the reasons for variations in treatments and outcomes across hospitals and areas. 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. They also allow much more precise analysis of practices at particular hospitals, providing a clearer evaluation mechanism for improving practices in the future.

Our method may be generally applicable for estimating treatment effects and for elucidating the underlying reasons for the large differences in health outcomes across providers. The method can, in principal, be applied to any outcome and any combination of treatments for a condition, as long as there is variation in treatment rates across hospitals, geographic areas, and other units of observation (e.g., physicians, medical groups, or health plans), and as long as selection bias across these observational units is not substantial. We have also developed some empirical tests for evaluating the extent to which selection bias is likely to be a problem. Evaluating the performance of our methods in other applications, in health care as well as other fields, is an important topic for future work.

References

- Allison JJ, Kiefe CI, Weissman NW, Person SD, Rousculp M, Canto JG, Bae S, Williams OD, Farmer R, Centor RM, 2000, "Relationship of hospital teaching status with quality of care and mortality for Medicare patients with acute MI," *JAMA*. 2000 Sep 13; 284(10):1256-62.
- Angrist, J.D. and A.B. Krueger, 1995, "Split-Sample Instrumental Variables estimates of the Return to schooling," *JBES*, April 1995, 13(2):225-35.
- Angrist, J.D., G.W. Imbens, and A.B. Krueger, 1999, "Jackknife Instrumental Variables Estimation" *Journal of Applied Econometrics*, Jan-Feb 1999, 14(1):57-67.
- Asch, S.M., E.M. Sloss, C. Hogan, R.H. Brook, R.L. Kravitz, 2000, "Measuring Underuse of Necessary Care Among Elderly Medicare Beneficiaries Using Inpatient and Outpatient Claims," *JAMA*, November 8, 2000, 284, 2325-.
- 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, A., 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.
- Hansen, L.P., J. Heaton and A. Yaron, 1996, "Finite Sample Properties of Some Alternative GMM Estimators" *JBES*, July 1996, 14(3):262-80.

- Johnson, R.A. and D.W. Wichern, 1998, *Applied Multivariate Statistical Analysis*, 4th ed., Upper Saddle River, NJ, Prentice Hall, 1998 (chapter 9).
- Marciniak, T., E. Ellerbeck, M.J. Radford., 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, M. and H. Noguchi, 2000, "Valid Estimation of Medical Treatment Effects Using Observational Data: Intensive Treatment of Heart Attacks in the Elderly," working paper, June.
- McClellan, M. and J. P. Newhouse, 1997, "The Marginal Cost-Effectiveness of Medical Technology: A Panel Instrumental-Variables Approach," *Journal of Econometrics*, 77(1), March 1997, 39-64.
- McClellan, M., and J.P. Newhouse, 2000, "Overview: Instrumental Variables Methods in Health Services Research," *Journal of Health Services Research*, forthcoming.
- McClellan, M. and D. Staiger, 1999a, "Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals," in D. Cutler (ed.) *Not-for-Profit Hospitals*, Chicago, University of Chicago Press.
- McClellan, M. and D. Staiger, 1999b, "Evaluating health care providers," in A. Garber, ed., *Frontiers in Health Policy Research, Vol. 3*, MIT Press, forthcoming.
- McClellan, M. and D. Staiger, 2000, "The Quality of Health Care Providers," Working Paper (revised from NBER Working Paper #7327, 1999).
- Normand, S.T, ME Glickman and C.A. Gatsonis, 1997, "Statistical Methods for Profiling Providers of Medical Care: Issues and Applications," *JASA*, 92(439), 1997, 803-814.
- Rosenbaum, P., and D. Rubin, 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 Immediate Percutaneous Angioplasty," *Lancet* 197-202.

- Staiger, D. and J. H. Stock, 1997, "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65(3), May 1997, 557-586.
- Stock, JH and J. Wright, 1996, "Asymptotics for GMM estimators with weak instruments" NBER technical working paper #198, July 1996.
- Thomas JW, Hofer TP, 1999, "Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care," *Med Care* 1999 Jan; 37(1):83-92.
- 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, "Outcomes in patients with acute non-Q-wave myocardial infarction randomly assigned to an invasive as compared with a conservative management strategy. Veterans Affairs Non-Q-Wave Infarction Strategies in Hospital (VANQWISH) Trial Investigators", *N Engl J Med* 1998 Jun 18;338(25):1785-92. Published erratum appears in *N Engl J Med* 1998 Oct 8;339(15):1091.
- Zijlstra, F., et al., 1993, "A Comparison of Immediate Coronary Angioplasty with Intravenous Streptokinase in Acute Myocardial Infarction," *New England Journal of Medicine* 328:680-684.

Table 1: Summary Hospital Statistics
(standard deviations in parentheses)

Admissions	
Mean Number of admissions	43.1 (46.5)
Median Number of admissions	27.0
Outcomes	
30-day mortality	17.2 (6.2)
1-year mortality	29.5 (7.0)
Treatments	
90-Day Catheterization	47.6 (14.1)
90-DAY CABG	16.0 (7.1)
1-DAY PTCA	8.7 (9.8)
90-Day PTCA	11.0 (6.2)
ACE Inhibitor at discharge	25.3 (8.4)
Thrombolytics in hospital	15.2 (8.6)
Aspirin in hospital	78.5 (10.9)
Beta Blockers in hospital	46.7 (14.6)
Chest X-ray	89.1 (9.2)
CK (Blood test for MI)	96.0 (5.7)
Blood glucose test	96.1 (5.4)
Hemoglobin count	96.4 (4.6)

Weighted means and standard deviations across hospitals were computed for a sample of 3963 hospitals with at least 3 admissions, with adjustment for a detailed set of patient covariates. Each hospital was weighted by the number of admissions.

**Table 2: GMM estimates of standard deviations and correlations among outcomes
Adjusted for detailed set of patient covariates**
(SDs on diagonal; correlations off diagonal; std. errors in parentheses)

	30-day mortality	1-year mortality
30-day mortality	0.032 (0.003)	0.901 (0.053)
1-year mortality	0.901 (0.053)	0.032 (0.003)

**Table 3: GMM estimates of standard deviations and correlations among treatments
Adjusted for detailed list of patient covariates**

(SDs on diagonal; correlations off diagonal; std. errors in parentheses)

	90-Day Cath	90-DAY CABG	1-DAY PTCA	90-Day PTCA	ACE Inhibitor at discharge	Thrombolytics in hospital	Aspirin in hospital	Beta Blockers in hospital	Chest X-ray
90-Day Cath	0.126 (0.002)	0.590 (0.033)	0.737 (0.010)	0.205 (0.027)	0.179 (0.033)	-0.314 (0.024)	0.321 (0.022)	0.026 (0.022)	0.434 (0.022)
90-DAY CABG	0.590 (0.033)	0.046 (0.002)	0.395 (0.031)	-0.082 (0.042)	0.068 (0.048)	-0.226 (0.037)	0.181 (0.032)	0.024 (0.030)	0.341 (0.034)
1-DAY PTCA	0.737 (0.010)	0.395 (0.031)	0.089 (0.001)	-0.241 (0.021)	0.312 (0.027)	-0.597 (0.017)	0.353 (0.015)	0.046 (0.02)	0.601 (0.019)
90-Day PTCA	0.205 (0.027)	-0.082 (0.042)	-0.241 (0.021)	0.042 (0.001)	-0.142 (0.043)	0.413 (0.034)	-0.045 (0.032)	-0.035 (0.03)	-0.307 (0.027)
ACE Inhibitor at discharge	0.179 (0.033)	0.068 (0.048)	0.312 (0.027)	-0.142 (0.043)	0.054 (0.002)	-0.203 (0.035)	0.531 (0.036)	0.454 (0.033)	0.218 (0.033)
Thrombolytics in hospital	-0.314 (0.024)	-0.226 (0.037)	-0.597 (0.017)	0.413 (0.034)	-0.203 (0.035)	0.069 (0.002)	-0.080 (0.027)	-0.029 (0.026)	-0.490 (0.022)
Aspirin in hospital	0.321 (0.022)	0.181 (0.032)	0.353 (0.015)	-0.045 (0.032)	0.531 (0.036)	-0.08 (0.027)	0.092 (0.002)	0.455 (0.02)	0.264 (0.025)
Beta Blockers in hospital	0.026 (0.022)	0.024 (0.030)	0.046 (0.020)	-0.035 (0.030)	0.454 (0.033)	-0.029 (0.026)	0.455 (0.020)	0.129 (0.002)	0.081 (0.022)
Chest X-ray	0.434 (0.022)	0.341 (0.034)	0.601 (0.019)	-0.307 (0.027)	0.218 (0.033)	-0.49 (0.022)	0.264 (0.025)	0.081 (0.022)	0.080 (0.002)

Table 4: Estimates of effect of catheterization on 30-day and 1-year mortality
Comparison of estimates from GMM, IV and OLS methods
with and without detailed patient covariates
(Standard errors of estimates in parentheses)

90-day catheterization effects	30-day mortality	1-year mortality	30-day mortality	1-year mortality
	Detailed Controls		Demographic Controls	
GMM	-8.0 (1.0)	-12.0 (1.1)	-8.0 (1.0)	-14.9 (1.1)
IV	-11.3 (3.4)	-12.2 (3.9)	-10.7 (3.5)	-10.9 (4.1)
OLS	-12.8 (0.2)	-15.8 (0.2)	-20.8 (0.2)	-28.7 (0.2)

Table 5A: IV Estimates of Hospital Practice Pattern Effects
Based on AMI admissions from the CCP Project, 1994-95

PRACTICE PATTERN/ FACTOR	30-DAY MORTALITY						Exp. Var.
	1	2	3	4	5	6	
<i>MORTALITY EFFECT</i>	-0.039 (0.107)	-1.549 (0.228)	-0.027 (0.123)	-0.669 (0.197)	-1.035 (0.272)	-0.234 (0.164)	0.390
90-DAY CATH	9.658 (0.466)	1.527 (0.251)	2.925 (0.301)	4.750 (0.515)	1.021 (0.378)	3.256 (0.294)	0.862
90-DAY CABG	1.189 (0.106)	0.282 (0.131)	0.689 (0.153)	4.403 (0.219)	0.323 (0.127)	-0.149 (0.108)	0.999
1-DAY PTCA	7.151 (0.401)	1.554 (0.143)	3.293 (0.307)	0.956 (0.213)	2.435 (0.347)	-1.330 (0.163)	0.910
90-DAY PTCA	0.199 (0.087)	-0.176 (0.108)	-0.977 (0.096)	-0.068 (0.096)	-0.633 (0.145)	3.995 (0.140)	0.999
ACE INHIBITOR AT DISCHARGE	0.831 (0.151)	3.588 (0.222)	0.608 (0.167)	-0.207 (0.210)	0.371 (0.233)	-0.480 (0.198)	0.506
THROMBOLYTICS IN HOSPITAL	-2.049 (0.498)	-0.419 (0.228)	-2.094 (0.191)	-0.313 (0.201)	-5.038 (1.332)	1.757 (0.421)	0.772
ASPIRIN IN HOSPITAL	2.352 (0.243)	6.577 (0.407)	1.113 (0.224)	0.594 (0.285)	-0.899 (0.460)	-0.124 (0.269)	0.582
BETA BLOCKERS IN HOSPITAL	-1.308 (0.273)	9.047 (0.499)	0.031 (0.286)	0.061 (0.316)	0.439 (0.784)	0.076 (0.314)	0.503
CHEST X-RAY	2.302 (0.250)	1.133 (0.208)	5.628 (0.338)	1.157 (0.251)	1.651 (0.387)	-0.983 (0.213)	0.686
CK (BLOOD TEST FOR MI)	0.427 (0.153)	1.106 (0.123)	3.132 (0.257)	1.132 (0.179)	1.323 (0.477)	-0.161 (0.109)	0.621
BLOOD GLUCOSE TEST	0.439 (0.114)	0.331 (0.109)	3.233 (0.197)	0.054 (0.113)	0.123 (0.101)	-0.310 (0.097)	0.544
HEMOGLOBIN COUNT	0.462 (0.087)	-0.241 (0.098)	2.712 (0.201)	-0.016 (0.102)	0.007 (0.200)	-0.385 (0.094)	0.601
Explained Hospital Variance	0.220	0.199	0.116	0.062	0.053	0.044	

Standard errors of estimates in parentheses. Factor effects and treatment loadings reported in percentage points; they show the effect of a one-unit (one standard deviation) change in the factor. Includes adjustment for demographics, co-morbidities, and severity of illness; results with demographic controls only do not differ substantially.

Table 5B: IV Estimates of Hospital Practice Pattern Effects
Based on AMI admissions from the CCP Project, 1994-95

PRACTICE PATTERN/ FACTOR	1-YEAR MORTALITY						Exp. Var.
	1	2	3	4	5	6	
<i>MORTALITY EFFECT</i>	-0.267 (0.187)	-1.495 (0.231)	-0.044 (0.157)	-0.966 (0.256)	-0.935 (0.423)	-0.350 (0.222)	0.390
90-DAY CATH	9.421 (0.525)	1.604 (0.305)	2.954 (0.313)	4.978 (0.598)	0.910 (0.844)	3.268 (0.305)	0.848
90-DAY CABG	1.131 (0.125)	0.252 (0.145)	0.672 (0.165)	4.425 (0.222)	0.260 (0.158)	-0.19 (0.118)	0.999
1-DAY PTCA	7.261 (0.628)	1.645 (0.150)	3.182 (0.362)	1.068 (0.256)	2.545 (0.727)	-1.329 (0.192)	0.934
90-DAY PTCA	0.167 (0.142)	-0.166 (0.107)	-0.961 (0.101)	-0.039 (0.099)	-0.592 (0.237)	4.008 (0.146)	0.999
ACE INHIBITOR AT DISCHARGE	0.787 (0.146)	3.632 (0.228)	0.591 (0.176)	-0.156 (0.212)	0.368 (0.278)	-0.486 (0.202)	0.515
THROMBOLYTICS IN HOSPITAL	-2.161 (0.725)	-0.431 (0.211)	-2.079 (0.252)	-0.408 (0.211)	-4.412 (1.386)	1.898 (0.518)	0.666
ASPIRIN IN HOSPITAL	2.291 (0.307)	6.583 (0.436)	1.146 (0.228)	0.678 (0.290)	-1.071 (0.575)	-0.135 (0.277)	0.595
BETA BLOCKERS IN HOSPITAL	-1.580 (0.271)	9.179 (0.543)	-0.069 (0.299)	0.165 (0.369)	0.899 (0.799)	0.099 (0.322)	0.521
CHEST X-RAY	2.260 (0.289)	1.171 (0.221)	5.532 (0.394)	1.237 (0.285)	1.840 (0.499)	-1.003 (0.234)	0.680
CK (BLOOD TEST FOR MI)	0.392 (0.187)	1.071 (0.124)	3.071 (0.303)	1.177 (0.195)	1.522 (0.633)	-0.178 (0.131)	0.629
BLOOD GLUCOSE TEST	0.476 (0.134)	0.34 (0.111)	3.208 (0.199)	0.066 (0.120)	0.132 (0.132)	-0.326 (0.099)	0.534
HEMOGLOBIN COUNT	0.456 (0.087)	-0.221 (0.099)	2.799 (0.256)	-0.012 (0.105)	-0.062 (0.243)	-0.378 (0.091)	0.648
Explained Hospital Variance	0.217	0.203	0.113	0.067	0.048	0.045	

Standard errors of estimates in parentheses. Factor effects and treatment loadings reported in percentage points; they show the effect of a one-unit (one standard deviation) change in the factor. Includes adjustment for demographics, co-morbidities, and severity of illness; results with demographic controls only do not differ substantially

Table 6: IV Estimates of Zip Code Practice Pattern Effects
Based on AMI admissions from the CCP Project, 1994-95

PRACTICE PATTERN/ FACTOR	30-DAY MORTALITY		Exp. Var.
	1	2	
<i>MORTALITY EFFECT</i>	-0.922 (0.196)	-0.144 (0.183)	0.399
90-DAY CATH	-0.808 (0.445)	5.175 (0.341)	0.821
90-DAY CABG	0.162 (0.205)	0.639 (0.249)	0.551
1-DAY PTCA	-0.678 (0.357)	3.661 (0.243)	0.903
90-DAY PTCA	0.140 (0.208)	0.311 (0.238)	0.681
ACE INHIBITOR AT DISCHARGE	1.166 (0.258)	0.237 (0.269)	0.402
THROMBOLYTICS IN HOSPITAL	0.711 (0.251)	-0.269 (0.251)	0.237
ASPIRIN IN HOSPITAL	1.576 (0.294)	1.836 (0.293)	0.370
BETA BLOCKERS IN HOSPITAL	7.248 (0.247)	-0.043 (0.651)	0.987
CK (BLOOD TEST FOR MI)	-0.008 (0.124)	-0.383 (0.151)	0.674
BLOOD GLUCOSE TEST	-0.592 (0.157)	-0.124 (0.159)	0.642
HEMOGLOBIN COUNT	-0.860 (0.14)	-0.063 (0.145)	0.897
Explained Hospital Variance	0.357	0.264	

Standard errors of estimates in parentheses. Factor effects and treatment loadings reported in percentage points; they show the effect of a one-unit (one standard deviation) change in the factor. Includes adjustment for demographics, co-morbidities, and severity of illness; results with demographic controls only do not differ substantially

Table 7: Goodness-of-Fit Test

Outcome Measure	Number of Factors	GMM Statistic	D.F.	P-value Overall	P-value versus one more factor
30-day mortality					
Hospital IV	5	130.1	23	0.000	0.000
	6	37.7	15	0.001	0.0003
	7	10.7	8	0.222	0.999
Zip IV	1	189.8	54	0.000	0.000
	2	69.9	43	0.006	0.0013
	3	40.9	33	0.163	0.103
1-year mortality					
Hospital IV	5	122.8	23	0.000	0.000
	6	39.6	15	0.001	0.0007
	7	14.3	8	0.075	0.720