A Comparison of Treatment Effects Estimators Using a Structural Model of AMI Treatment Choices and Severity of Illness Information from Hospital Charts*

Ahmed Khwaja
Fuqua School of Business
Duke University†

Gabriel Picone
Department of Economics
University of South Florida

Martin Salm
Department of Econometrics
Tilburg University

Justin G. Trogdon
RTI International

Abstract

We compare the performance of various matching estimators using a novel approach that is feasible in the absence of experimental data. We estimate a structural model of hospital choices and catheterization for Medicare heart attack victims using hospital chart data on patient heterogeneity. With the estimated structural parameters, we simulate data for which the treatment effect is known. We find that as measures of individual heterogeneity are added to the controls, matching estimators perform well. However, the estimators do a poor job recovering the true treatment effect when measures of individual heterogeneity are unavailable.

KEYWORDS: Matching estimators, treatment effects, catheterization, dynamic discrete choice estimation

JEL CLASSIFICATION: I12, C31, C35

*We would like to thank the editor, John Rust, and four anonymous referees for their suggestions. We have also benefited from comments of Ron Gallant, Han Hong, Jeff McCullough, Paul Miller, Frank Sloan, Andrew Jones, and seminar participants at University of Adelaide, University of York, and the ASHE Biennial Conference 2008. This research was supported by a grant from the Agency for Healthcare Research and Quality, "Hospital Volume and Quality of Care," (1R03-HS013386-01A1). We alone are responsible for any remaining errors. The contents of this publication do not necessarily reflect the views of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. The authors assume full responsibility for the accuracy and completeness of the ideas presented. This article is a direct result of the Health Care Quality Improvement Program initiated by the U.S. Health Care Financing Administration, which has encouraged identification of quality improvement projects derived from analysis of patterns of care.

†1 Towerview Drive, Durham, NC 27708. Phone: 919-660-1989. Fax: 919-660-7971.

1 Introduction

Estimating the causal effect of an intervention or treatment in the presence of unobserved heterogeneity is a common problem in economics. The literature on program evaluation has suggested several ways to estimate the effects of social programs on individual labor market and other outcomes (see e.g., Angrist and Krueger 1999, Heckman, LaLonde and Smith 1999). Each of these estimators requires a different set of assumptions and may even estimate different effects (see e.g. Imbens 2004, Cobb-Clark and Crossley 2003). This has led to a lack of consensus about the merits of these estimators. In this paper, we compare the performance of a number of these estimators by estimating the effects of an important medical intervention, catheterization, on survival outcomes of individuals with heart attacks, i.e. acute myocardial infarction (AMI). For AMI patients, severity of illness is a primary factor for determining treatment and if omitted leads to bias in the treatment-outcome relationship. In the *absence* of a measure of the *true* treatment effect we adopt a novel approach. We estimate a structural model of treatment and hospital choices for AMI patients using a unique data set that combines administrative Medicare claims data with hospital chart level data on patient severity of illness. With the estimated structural model we simulate data for which the treatment effect is known to us, and we use this data to compare various treatment effect estimators.

A treatment effect estimator attempts to measure the difference in outcomes for an individual between being treated and not being treated. This treatment effect may not be the same for all patients. For example, the benefit of an invasive procedure like catheterization will not be the same for a 55 year old, relatively healthy patient and a 95 year old, frail patient. Thus, it is important to first determine the population for which one would like to obtain consistent estimates of the treatment effects. The most common treatment effect estimated in the literature is the average treatment effect, which is the effect of treatment on a randomly drawn person from the general population. However, many individuals will never receive treatment (e.g., many patients will never be candidates for catheterization). In these instances a more interesting treatment effect is the average treatment effect on the treated, which is the average gain due to the treatment for those individuals who actually received treatment. There are other treatment effects, like the marginal treatment effect, that have been studied in the literature, but will not be part of this study (e.g., see Heckman and Vytaclil 2005).

We compare different treatment estimators with respect to their ability to estimate the effects of catheterization on individuals with AMI on their one year mortality outcome. We estimate both the average treatment effect and the average treatment effect on the treated. Obtaining reliable causal estimates of this treatment effect has important policy implications and has received prominent attention in the past. Moreover, researchers have estimated the effect of catheterization on mortality using standard administrative Medicare claims data, e.g., McClellan et al. (1994) and McClellan and Newhouse (1997).[1]

One significant limitation of Medicare claims data is that it does not contain comprehensive measures of severity of illness. Unmeasured severity of illness is an example of unobserved heterogeneity that leads to omitted variable bias in the estimation of the treatment-outcome relationship, e.g., patients that receive catheterization may be healthier than patients who do not. Researchers have tried to overcome this problem using instrumental variables. Following the work of McClellan et al. (1994) on the average treatment effects of undergoing catheterization, which demonstrated that the differential distance between the nearest catheterization hospital and nearest hospital was correlated with the probability of receiving treatment and uncorrelated with measurable indicators of quality, the differential distance has become a widely used instrument to study different treatment effects for AMI patients. For example, Gowrisankaran and Town (1999) employ the differential distance between the nearest for-profit hospital and nearest hospital as an instrument for admission to a for-profit hospital.

The strategy of this study is to use a richer data set to estimate a structural model of treatment and hospital choices. The estimated model is then used as a data generating process to simulate a data set in which the treatment effect is known to us. The simulated sample, which includes variables from hospital chart data, is used to analyze the sensitivity of different estimators to the addition of detailed severity of illness measures. In estimating the model, we use data from the Cooperative Cardiovascular Project (CCP) merged with the American Hospital Association's (AHA) annual survey of hospitals and Medicare claims data. The CCP is a sample of over 180,000 Medicare patients who were admitted for AMI during 1994-95. The CCP data is unique because it contains extremely detailed measures of severity of illness obtained from medical chart abstractions. It allows us to obtain the best possible controls for severity of illness and comorbidities for AMI patients (Krumholz et al. 2001). Another advantage of the CCP is that it contains the patient's zip code, which allows us to create the distance instrumental variable widely used in studies using instrumental variables and Medicare claims data.

[1]Other research has estimated the effects of, e.g., admission to for-profit hospitals (e.g., Gowrisankaran and Town 1999, Sloan et al. 2001) or admission to high-volume hospitals (e.g., Luft et al. 1990, Hamilton and Hamilton 1998, and Ho 2002) using different outcomes, e.g., 30 day mortality.

A number of influential studies in the evaluation of social programs literature have compared the performance of different estimators for treatment effects based on the assumption of ignorability of treatment using experimental and non-experimental data (e.g., LaLonde 1986; Dehejia and Wabha 1999; Heckman, Ichimura, and Todd 1997, 1998; Michalopoulos, Bloom, and Hill 2002). These studies conclude that the capability of non-experimental estimators to replicate experimental results is mixed. It appears that the performance is better when comparing treatment and control groups that are similar in geographic location and other observable characteristics. The goal of our study is similar to these analyses but there are important differences. The combination of experimental and non-experimental data of the type used in the above mentioned studies is rarely available and hence this type of evaluation cannot be conducted. Some studies in the medical literature have resorted to randomized clinical trials, but these studies are costly and the sample sizes are small relative to natural experiments. Furthermore, for many applications of interest, e.g., catheterization or admissions to for-profit hospitals, randomized clinical trials are infeasible. Hence in comparing estimators we use simulated data where the true treatment effect is known to us, in combination with information on detailed measures of severity of illness, the lack of which in any case is the main reason for omitted variable bias.

The focus of our analysis is to recover ATE and ATT using matching estimators although we also consider instrumental variable methods. We first show that we are able to recover the parameters from the structural model when the researcher has information on detailed patient characteristics. However, if either detailed patient characteristics or hospital characteristics are not observed, we are not able to recover the parameters from the structural model. We find that as measures of heterogeneity like patient characteristics are added to the regression specification matching estimators perform well in recovering the "true" effect (ATE and ATT). However, the estimators do a poor job in recovering the true treatment effect when the data is poor in such measures of individual heterogeneity. Our results suggest that users of these estimators should exercise caution as these estimators are not substitutes for data rich in measures of individual heterogeneity.

The rest of the paper is organized as follows. Section 2 describes the data, estimation sample, the structural model, and the procedure used to simulate the data that is used to evaluate the estimators. In Section 3 we provide an overview of the estimators we evaluate. Section 4 describes the alternative empirical specifications adopted in estimating the treatment effects regressions, and the variables used in these regressions. Section 5 presents the results of the comparison of different treatment effects estimators and section 6 concludes.

2 Dynamic Model of Hospital and Treatment Decisions

In the absence of an empirical (e.g., experimental) measure of the true treatment effect of catheterization on one year mortality we adopt the following strategy to compare estimators. We estimate a dynamic structural model of hospital and treatment choices and the consequent mortality outcomes. The estimated model is used to simulate data with known treatment effects. The performance of the estimators under consideration is then compared using the simulated data under various specifications.

The process of seeking care after an individual suffers an AMI is quite complex. While abstracting from various details, the model accounts for important features: (i) the sequential nature of the process, (ii) the effect of information revealed during the process on current decisions, and (iii) the effect of current decisions on the choice alternatives available in future.[2] We first discuss the data used to estimate the model and then describe the model and its empirical specification.

2.1 Data and Estimation Sample

The primary source of data is the Cooperative Cardiovascular Project (CCP). The CCP sample consists of randomly selected patient records for Medicare patients admitted to nonfederal acute-care hospitals in the US with a primary diagnosis of acute myocardial infarction (AMI), or heart attack, (ICD-9-CM 410, excluding a fifth digit of 2) over an eight-month period. All of the sampling occurred between February 1994 and July 1995. The sample includes all hospitals in the US that had not participated in a four-state pilot phase (Alabama, Connecticut, Iowa, and Wisconsin). Medical records for each sampled hospitalization were forwarded to clinical abstraction centers. Altogether, charts were abstracted for approximately 180,000 AMI patients. The CCP took steps to ensure consistent data collection including comparison of randomly selected records across two independent data abstraction centers. The results indicate a high degree of consistency in data abstraction. See Marciniak et al. (1998) for a detailed description of CCP goals, sampling, and data collection methods.

Table 1 about here

[2] In the model individuals and their providers are considered to have identical preferences over hospitals and treatments. The model is described in detail in Trogdon (2009).

We complement the CCP data set with hospital variables from two sources. First, we obtain information on availability of facilities for cardiac catheterization, angioplasty, and open-heart surgery at the hospital from the American Hospital Association's (AHA) annual survey of hospitals for different years. Second, we calculate annual hospital heart surgery volume for Medicare patients from within Medicare claims merged to the CCP.

The model is estimated for a sample from the CCP based on the following criteria. First, since the CCP is a sample of Medicare beneficiaries, each admission should have Part A claims data available; individuals who have admissions without such data, or admissions to other specialized types of hospitals, are dropped (N = 8,565).

Second, in order to make sure that the complete episode, or sequence of admissions, for a particular heart attack is available, all individuals are dropped who have previous episodes in the data or are first admitted from somewhere other than home (N = 47,432). This sample cut is made for two reasons. First, patients initially transferred from another hospital have had previous and unknown care. Second, the distance calculations are made based on zip code of residence. For patients admitted from other locales (e.g., nursing homes), these calculations would be incorrect.

Third, individuals who choose hospitals outside of the designated choice set are dropped (N = 11,139). [3] The choice set is the five nearest hospitals in each of two service categories, no specialized services and catheterization only, and the 10 nearest hospitals with open heart surgery capabilities. Each of these service categories provides a unique set of treatment choices. At the same time, heart attack victims are not expected to bypass several hospitals for treatment. This choice set captures 93% of the observed hospital choices in the CCP. In addition, this definition of the choice set keeps the size of the state and parameter spaces manageable. The remaining sample of patients is more likely to have had their heart attacks close to home, minimizing the error in the distance calculations.

Finally, patients who have unexplained sequences of procedures are dropped from the sample (N = 5,053). For example, these include patients with procedures at hospitals that do not have a record for those facilities, patients who have surgery recorded without a diagnostic procedure, patients who have multiple procedures recorded, and a small set of patients who transfer to hospitals that do not offer heart surgery. This leaves a sample of 114,818 patients for the analysis.

[3] Individuals who live in Alaska are also dropped from the analysis set. The distances in the state are non-representative of the rest of the sample.

Table 1 provides summary statistics for the sample from the CCP data. All individuals in the sample are older than 65, with 16.6% of the sample older than 85. There are approximately the same number of males as females, while about 10% are non white. About 74% of the sample resides in urban areas. There is substantial variation in health status and severity of heart attacks. Nearly half of the sample has important comorbidities as measured by the Charlson index (see Charlson et al. 1987 for details). Higher values of the Charlson index indicate worse health.

An important measure that provides an initial assessment of the severity of the heart attack at the time of admission is Killip class. Using a method developed by Killip and Kimbal (1967), heart attack patients are classified into one of four classes: (i) Class I: those with no evidence of congestive heart failure (CHF), (ii) Class II: those with mild to moderate CHF, (iii) Class III: those with overt pulmonary edema, (iv) Class IV: and those in cardiogenic shock. Thus, a higher classification indicates a more severe heart attack. Approximately half of the sample has evidence of at least moderate CHF at presentation to the hospital (i.e., Killip class greater than I).

A third measure of severity that we use is blockage status. This is measured using a discrete variable that corresponds to three blockage status categories: (1) normal (left ventricular ejection fraction $\geq$ 60%), (2) mild to moderate reduction (30 to 59%), and (3) severe reduction in left ventricular ejection fraction (0 to 29%). The vast majority of those patients who receive information concerning systolic function by having catheterization show a moderate reduction; approximately 10% of individuals have severe reduction.

An indicator variable, "Service," designates one of three possibilities for each hospital and is coded as: "1 = neither catheterization nor surgery services available," "2= catheterization available but not surgery," "3 = both catheterization and surgery available." A second indicator variable accounts for the annual "Volume" of surgeries and takes on one of four values: "1 = 0-50 surgeries," "2 = 51 to 100 surgeries," "3 = 101 to 200 surgeries,"and "4 = 200+ surgeries." Most individuals choose no-service hospitals for the initial admission. These are the most common type of hospital in the sample and, on average, are closest to the individual. Catheterization-only hospitals are the rarest type of hospital in the sample.

Treatment choices have two dimensions; a combination of a choice for procedure and transfer to another hospital if the procedure is not available at the current hospital. In the sample, 34.8% of the individuals have catheterization and half of those patients go on to have surgery. Approximately 15%

of patients are transferred during the treatment process. One year after admission, one third of the individuals in the sample have died. We next describe the dynamic model in detail.

2.2 Timing

There are four periods in the dynamic model which adopts a random utility framework (McFadden 1981).[4] In the first period, individuals have information about initial health status and preferences for types of hospitals. They suffer a heart attack and draws a stochastic component of preferences for types of hospitals (described in more detail below in Section 2.6), which is unobserved to the econometrician. Individuals subsequently choose a hospital. In the second period, individuals acquire initial information about the severity of the heart attack. They also draw a stochastic component of preferences to undergo catheterization. After this the individuals chooses whether or not to undergo catheterization, which could include transfer to another hospital. In the third period, conditional on choosing catheterization, individuals receive information about the systolic functioning of their heart. Then they draw a stochastic component of preferences to undergo surgery. Next they choose a surgery option, which again could include a transfer. Mortality outcomes are determined in the fourth and last period. Individuals do not make a decision in the last period but forms expectations about the outcomes that factor into decisions made earlier.

In addition to choice-specific random components of utility unobserved to the econometrician, and assumed uncorrelated across decisions, the model includes time invariant unobserved heterogeneity at the individual level that affects both the preferences for hospitals and the Killip transition, thus allowing for correlation between treatment choices and health outcomes.

2.3 Choices

Individuals make choices over the type of hospital for admission, and whether to undergo catheterization and surgery, including the possibility of transfer. In the first period, individuals choose a type of hospital. Hospitals are characterized by the specialized services they offer and distance from the individual. Let $j = 1, 2, \ldots, J$ represent each hospital in the choice set and d_1 represents the choice made by the individual in period 1, $d_1 \in D_1 = \{1, ..., J\}$. We denote the choice of the j'th element of D_1 by d_1^j. Empirically, the choice set includes the five nearest no-service, five nearest catheterization-only

[4] See e.g., Khwaja (2001) and Arcidiacono et al (2007) for similar applications of sequential dynamic programming models.

hospitals, and the 10 nearest surgery hospitals ($J = 20$).

In the second period, individuals choose whether or not to undergo catheterization, which could include transfer to another hospital. Using the same notation as above, we denote the diagnosis choices in the second period by $c = 1, \ldots, C$ and define $d_2 \in D_2 = \{1, ..., C\}$. We denote the choice of the c'th element of D_2 by d_2^c. Empirically, the mutually exclusive choices include no transfer/no catheterization (d_2^1), no transfer/catheterization (d_2^2), transfer/no catheterization (d_2^3), and transfer/catheterization (d_2^4), respectively ($C = 4$). The choice set in the second period is restricted in some cases by the choice made in period one, e.g., if a no-service hospital is chosen in the first period, then the second period choices are limited to c =1, 3, or 4; no transfer/catheterization is not an option.

Finally, the mutually exclusive surgery options $s = 1, \ldots, S$ available to individuals in the third period are no transfer/no surgery (d_3^1), no transfer/surgery (d_3^2), transfer/no surgery (d_3^3), and transfer/surgery (d_3^4): $d_3 \in D_3 = \{1, ..., S\}$. As before, we denote the choice of the s'th element of D_3 by d_3^s. Again, the surgery choice set is conditional on previous decisions. Specifically, surgery is only available to individuals who received catheterization in the second period.[5]

2.4 State Variables

The hospital and treatment decisions are made using the information available at the time of the decision, which is defined by a collection of state variables. To ease the computational burden all state variables are discretized. The state variables describing individuals at the initial stage include demographic characteristics—age (65 to 74, 75 to 84, 85 and older); gender (male, female); race (white, minority); residence in a Metropolitan Statistical Area (MSA); and an index of health status. The Charlson index (Charlson et al 1987), $h \in H = \{1, 2, 3\}$, summarizes individuals' health status at the time of the heart attack. The availability of information concerning comorbidities at the time of the choice of hospital allows preferences for hospitals to vary by health status. Individuals also belong to a discrete health type $t \in T$ that is known to them but unobserved to the econometrician.

Each individual has a *unique* hospital choice set depending on the residential location of the individual (which adds to the computation burden in solving the model and hence affects our simulation

[5]Patients can be transferred and not have a procedure at the new hospital. Patients are transferred for a variety of reasons including the option of catheterization or surgery as well as for other procedures, diagnostic tests or hospital facilities. The model focuses on the decision to undergo catheterization and surgery, and transfers are an important decision path to consider. However, other decisions after transfer are not modeled further but would be captured by the (dis)utility of transfer.

strategy described below in Section 2.10). Hospital characteristics include specialized services (no specialized services, catheterization only, and surgery capabilities); distance from the individual in miles (10 categories defined by the deciles of the distance to the nearest hospital, specific distance categories are presented in Table 3); and Medicare surgery volume (zero to 50, 51 to 100, 101 to 200, and 201 and over). Let D represent the individual's demographic characteristics, X represents the hospital characteristics in the choice set, and Z represents all of the individual's stationary characteristics: $Z = (D, h, t, X)$.

Additional state variables include measures of severity of heart attack. These drive the dynamic trade offs individuals face in making hospital and treatment decisions. After the hospital choice, information is revealed concerning the severity of the heart attack. The information about severity of the heart attack is unique to our clinical data and not typically available in administrative discharge records. It provides important information to identify treatment choices. At admission the individual receives an initial severity diagnosis by categorization into a Killip class, $k \in K = \{1, 2, 3\}$. In the third period, the catheterization reveals information about the extent of blockage $b \in B = \{1, 2, 3\}$, where the categories correspond to normal, mild to moderate reduction, and severe reduction in left ventricular ejection fraction.

The information set at time n, I_n, grows each period as individuals learn more about the nature of the heart attack:

$$I_1 = (Z) \ , \tag{2.1}$$

$$I_2 = (d_1, k; Z) \ , \tag{2.2}$$

$$I_3 = (d_1, k, d_2, b \cdot (d_2^2 + d_2^4); Z) \ , \tag{2.3}$$

$$I_4 = (d_1, k, d_2, b \cdot (d_2^2 + d_2^4), d_3; Z) \ . \tag{2.4}$$

2.5 Transition Probabilities

Individuals are forward-looking and form expectations about the possible health states they will be in conditional on the decisions they make. In forming these expectations the individuals use the following transition probabilities.

The probability of Killip class k, $Pr(k|D, h, t; \delta)$, is a function of the information set in period one and takes a multinomial logit form, where δ is a vector of parameters. Age, gender, and race are included to account for differences in the severity of heart attacks across demographic groups. The probability of a particular Killip class is also determined by initial health status: sicker patients initially (e.g., patients with hypertension or previous heart attack) are likely to have a higher probability of a more severe heart attack. The Killip transition also includes (unobserved) health type-specific intercepts.

The probability of a particular blockage category b, $Pr(b|k(t), h, D; \beta)$ depends on the characteristics of the individual, including realized health status from the first period, and takes a multinomial logit form, where β is a vector of blockage parameters. The motivation for including demographic characteristics and initial health status is the same as in the Killip transition. Here, not only does initial health affect the severity of the heart attack, but also the initial severity assessment.

In the final period, nature determines whether the individual lives or dies. In this study we focus on mortality at one year post admission. In doing so, the model is able to trace the impact of patient characteristics, including health status, on the choice of hospital, the treatment process, and the mortality outcome. The probability of mortality in period four, $Pr(m|d_3, b, k(t), h, D, X; \eta)$, is given by the logit probability, where η is a vector of parameters. Similar to the other health transitions, mortality rates are allowed to differ across demographic groups. Previous health status $(b, k(t), h)$ is also an important determinant of mortality. The coefficient on the surgery decision, d_3, measures the effect of the treatment on mortality outcomes. This effect is also allowed to vary by surgery volume. This is where the benefit of high hospital surgery volume enters the patient's decision-making process.

The parameter estimates for surgery would be biased upward if unobservably healthy patients get surgery, either because hospitals *choose* to operate on healthier patients or patients with better expected outcomes *choose* to have surgery. The health status variables control for this type of selection. The model also accounts for unobserved heterogeneity to control for selection. The model specification allows unobserved health to directly affect the Killip transition and to indirectly affect the blockage and mortality transitions through its impact on Killip class.

2.6 Utility Functions

There is a utility function associated with choices in each period. Period one utility from choosing hospital j is

$$U_1^j(Z, \epsilon_1^j) = \alpha_{10}^j + \alpha_{11}^j h + \alpha_{12}^j t + \alpha_{13} dist + \alpha_{14} dist \cdot MSA + \epsilon_1^j \ . \tag{2.5}$$

Here α_{10}^j captures the baseline utility from a hospital of service type j. Utility in period one is also a function of health status, both observed (α_{11}^j) and unobserved (α_{12}^j), and the distance to the type of hospital chosen (α_{13}). Distance is interacted with a urban/rural indicator to make distance a better proxy for true travel costs (α_{14}); distance in a metropolitan area will have a longer actual travel time than the same distance in a rural area. The stochastic component of utility from hospital choice j, ϵ_1^j, represents factors affecting hospital choice that are known to the individual but are unobserved to the econometrician.

In period two, utility from the catheterization choice c varies by health status:

$$U_2^c(k, \epsilon_2^c) = \alpha_{20}^c + \alpha_{21}^c k + \epsilon_2^c \ . \tag{2.6}$$

The utility from surgery choice s in period 3 is similar, but individuals now have more information about severity (b):

$$U_3^s(b, \epsilon_3^s) = \alpha_{30}^s + \alpha_{31}^s b + \epsilon_3^s \ . \tag{2.7}$$

Catheterization and surgery impact the utility directly for two reasons. First, both of these invasive procedures involve pain and discomfort, especially in the recovery process. This is captured in the choice-specific intercepts α_{20}^c and α_{30}^s. Second, these procedures involve risks for side effects and these risks differ by health status. The same is true for transfer; transfer to another hospital is stressful and disorienting. Thus, there is a trade-off between the disutility and costs associated with the treatment choices and the benefits of the diagnostic information gained and the consequent improved outcomes through treatment. The ϵ_2^c and ϵ_3^s are the stochastic components of utility for catheterization and surgery, respectively, and represent factors affecting these choices that are known to the individual but are not observed by the econometrician.

2.7 Dynamic Optimization

In each period, individuals choose an option from the relevant choice set to maximize expected lifetime utility. In this model, the information available at the time the individual makes his decisions depends on his past decisions. The dynamic model can be solved recursively, working backwards from the last period. Since the choice set is unique to each individual, depending on the individual's residential location, the model has to be solved separately for each individual.[6]

In period three, individuals choose a surgery option to maximize the current utility from surgery and the expected value of utility next period. Terminal utility is normalized to zero for death and one for surviving the episode. Let $\overline{U}$ denote the deterministic part of utility. Then the expected lifetime utility from choosing s in period three is

$$V_3^s\,(I_3, \epsilon_3) = \overline{U}_3^s\,(b) + \epsilon_3^s + \mu\,(1 - Pr(m|\cdot))\ , \tag{2.8}$$

where μ is the discount factor.

In order to clearly show the effect of past choices on current and future choice sets, redefine $d_1 \in D_1 = \{1, 2, 3\}$, where $j = 1, 2, 3$ represent hospitals with no specialized services, catheterization only, and surgery services, respectively. The expected maximized lifetime utility from choosing c in period two is

$$\begin{aligned} V_2^c(I_2, \epsilon_2^c) = \overline{U_2^c}(k) + \epsilon_2^c + \\ \mu\Big\{\sum_{b=1}^{3} Pr(b|\cdot) \cdot \{E[max_{s\in\{1,3,4\}} V_3^s(I_3, \epsilon_3^s)] \cdot d_2^2 \cdot d_1^2 + \\ E[max_{s\in\{1,2\}} V_3^s(I_3, \epsilon_3^s)] \cdot (d_2^4 + d_2^2 \cdot d_1^3) + \\ E[V_3^1(I_3, \epsilon_3^1)] \cdot (d_2^1 + d_2^3)\}\Big\}\ . \end{aligned} \tag{2.9}$$

The first two components of (2.9) are just the utility from choice c. The expected maximized utility in period two has three components. The first is the expected utility next period when choosing no transfer/catheterization at a catheterization-only hospital in period two. The expectations are with respect to ϵ_3. In this situation, the surgery options are no transfer/no surgery, transfer/no surgery, and transfer/surgery. The second component is the expected utility next period when catheterization is

[6] An alternative would have been to assume the same choice set for everyone from the universe of hospitals in our data but this would have led to an extremely large choice set, and also seemed a less plausible assumption.

chosen at a surgery hospital, either by transferring to a surgery hospital (d_2^4) or by choosing a surgery hospital in the first period ($d_2^2 \cdot d_1^3$). In this situation, transfer is not necessary. The final component is the expected utility next period when choosing no catheterization, which leaves only the no transfer/no surgery option. [7]

The expected maximized lifetime utility from choosing hospital j in period one is

$$\begin{aligned} V_1^j(I_1, \epsilon_1^j) = \overline{U_1^j}(Z) + \epsilon_1^j + \\ \mu \Big\{ \sum_{k=1}^{3} Pr(k|\cdot) \cdot \{ E[max_{c\in\{1,3,4\}} V_2^c(I_2, \epsilon_2^c)] \cdot d_1^1 + \\ E[max_{c\in\{1,2,3,4\}} V_2^c(I_2, \epsilon_2^c)] \cdot d_1^2 + \\ E[max_{c\in\{1,2\}} V_2^c(I_2, \epsilon_2^c)] \cdot d_1^3 \} \Big\} . \end{aligned} \tag{2.10}$$

Again, each choice of hospital corresponds to a different choice set available next period. Individuals that choose a hospital without specialized services (d_1^1) cannot choose to have catheterization without transfer in the next period. Individuals that choose a surgery hospital (d_1^3) do not have to transfer to another hospital in future periods.

2.8 Estimation

The unobserved taste parameters, ϵ_n, are assumed to be i.i.d. Type I Extreme Value $\left(\rho_n \gamma, \frac{\pi^2 \rho_n^2}{6}\right)$, where γ is the Euler constant and ρ_n is the scale parameter for period n. This distributional assumption allows for estimation of the dynamic model by yielding a closed-form solution for the expected value of the future value functions and choice probabilities, conditional on unobserved health type t. Specifically, let $\overline{V_n^l}$ represent the deterministic portion of the value function for choice l in period n. Then the probability of observing an individual making choice l in time n conditional on type t is

$$Pr(d_n^l | I_n) = \frac{exp\left(\overline{V_n^l}(I_n)/\rho_n\right)}{\sum_{l'=1}^{M} exp\left(\overline{V_n^{l'}}(I_n)/\rho_n\right)} . \tag{2.11}$$

The health type t is unobserved to the econometrician. If an individual does not choose catheterization, neither he nor the econometrician know the blockage status. If an individual chooses

[7] Some patients in the sample are observed to have transferred hospitals but not to have received either catheterization or surgery at either hospital. For these cases it is assumed that the transfer was in period two; no transfer/no catheterization and transfer/no surgery is not observed in the data.

catheterization after transferring to a non-CCP hospital, then he, but not the econometrician, knows the blockage status. In all cases, individual choices provide information for the surgery and mortality probabilities. The likelihood of observing an individual's choices and health transitions when blockage is *observed* in the data integrates over the unobserved health type:

$$\begin{aligned} L_i^b = & \left\{ \sum_{t=1}^{T} Pr(t) \cdot \left(\prod_{j=1}^{J} Pr(d_1^j | I_1)^{d_1^j} \cdot \prod_{k=1}^{K} Pr(k|\cdot)^{1(k)} \right) \right\} \cdot \prod_{c=1}^{C} Pr(d_2^c | I_2)^{d_2^c} \cdot \\ & \prod_{b=1}^{B} Pr(b|\cdot)^{1(b)} \cdot \prod_{s=1}^{S} Pr(d_3^s | I_3)^{d_3^s} \cdot Pr(m|\cdot)^{1(m=1)} \left(1 - Pr(m|\cdot)\right)^{1(m=0)} . \end{aligned} \tag{2.12}$$

The unobserved type does not directly enter the health transitions and choices after the Killip transition; the effects are captured in the hospital choice and Killip state variables. Therefore, the likelihood function does not have to integrate over types for these probabilities.

To take advantage of the information concerning the surgery decision and mortality outcomes from all observations, the likelihood for individuals whose blockage was *unobserved* in the data integrates over the unobserved blockage status:

$$\begin{aligned} L_i^{\mathbf{no}b} = & \left\{ \sum_{t=1}^{T} Pr(t) \cdot \left(\prod_{j=1}^{J} Pr(d_1^j | I_1)^{d_1^j} \cdot \prod_{k=1}^{K} Pr(k|\cdot)^{1(k)} \right) \right\} \cdot \prod_{c=1}^{C} Pr(d_2^c | I_2)^{d_2^c} \cdot \\ & \left\{ \sum_{b=1}^{B} Pr(b|\cdot)^{1(b)} \cdot \left(\prod_{s=1}^{S} Pr(d_3^s | I_3)^{d_3^s} \cdot Pr(m|\cdot)^{1(m=1)} \left(1 - Pr(m|\cdot)\right)^{1(m=0)} \right) \right\} . \end{aligned} \tag{2.13}$$

Allowing for unobserved heterogeneity using a mixing distribution (e.g., Heckman and Singer 1984) removes the additive separability of part of the likelihood function, but estimation can still proceed in stages using the EM algorithm (Rust and Phelan 1997; Arcidiacono and Jones 2003). The inclusion of unobserved types produces correlation of the unobserved taste parameters with health states. It also allows for unobserved heterogeneity to affect assignments into health classes (i.e., heterogeneity in the way hospitals assign Killip class). The assumption of conditional independence of the unobserved taste parameters leads to a log likelihood function that is additively separable for all but the likelihood components of the hospital choices and Killip transitions.

First, the blockage transition is estimated as a standard multinomial logit.[8] The estimation of the utility parameters requires a nested algorithm which solves the dynamic model for each iteration

[8]The blockage parameters are estimated on the sub-sample of patients with observed blockage data (N = 27,390).

of the parameter values. Using estimates for the blockage transition (β), consistent estimates of the mortality and period three utility parameters (η and α_3) are obtained. Using these estimates, utility parameters for period two (α_2) are recovered from a multinomial logit for the catheterization decision. Finally, the EM algorithm is used to estimate the Killip transition (δ), period one utility parameters (α_1), and the type probabilities ($Pr(t)$). The number of discrete types (T) is fixed at two. Consistent standard errors are obtained by taking one Newton step on the full likelihood function at the consistent estimates from the multi-stage estimation (Rust 1994, p. 3109) and using the outer product of the gradient formula.

2.9 Parameter Estimates

The estimates of the model parameters are presented in Tables 2-4. The model is estimated with 2 unobserved types. The estimated probability of being of health type 2 is 9.4%. In general the parameters of the health transition probabilities are generally significant and have plausible signs (Table 2). To illustrate, the parameters of the Killip transition affect the probability of a particular Killip class in expected ways. In particular, older individuals and those in a higher Charlson category are more likely to transition to a higher Killip class. Relative to those of type 1, type 2 individuals are less likely to be of Killip class III or IV. Similar trends are found in the estimates for the blockage category transitions. Older individuals have a higher probability of transitioning into a more severe reduction category relative to the normal category. Also, lower prior health status raises the probability of worse systolic function, e.g., individuals in a higher Killip class have a higher probability of having a severe reduction (relative to normal function).

In the model, health status measures associated with every period affect one-year mortality (Table 2). The estimates imply that individuals with a Charlson index greater than 3 are 85% more like to die within a year of admission than those in the lowest Charlson category; individuals in Killip class III or IV are twice as likely to die than those with Killip class I; and individuals with severe reduction are three and a half times more likely to die than those with normal systolic function. The likelihood of death increases with age but there are no statistically significant differences by gender. Surgery reduces the likelihood of mortality, especially at higher surgery volume hospitals. This estimate controls for detailed health status, which controls for selection at surgery hospitals.

Table 2 about here

Table 3 about here

Table 4 about here

The estimates of the utility parameters for the hospital choice in period 1 are presented in Table 3. These parameters estimates are relative to choosing a hospital with no specialized services. Type 1 individuals prefer no-service hospitals, while those of type 2 prefer hospitals with catheterization and surgery services to no-service hospitals, all else constant. These results along with those for the Killip transition, imply that type 2 individuals are healthier and prefer hospitals with more services. Individuals with worse initial health status, as measured by the index of comorbidities, have a greater preference for catheterization and surgery hospitals relative to no-service hospitals suggesting that hospital choice is affected by health status. Individuals prefer to choose hospitals that are closer and the disutility from distance to hospital is greater for individuals living in urban areas due to the fact that travel time is higher in urban areas for a given distance.

The estimates of the utility parameters for the catheterization and surgery choices in periods 2 and 3 are presented in Table 4. These parameter estimates are relative to choosing the "no transfer/no procedure" option. The utility parameter estimates for catheterization imply that there is disutility from undergoing this procedure. The choice intercepts measure the utility from each choice relative to choosing no transfer and no catheterization and all three are negative and significant. The disutility from transfer and catheterization is probably due to pain and suffering associated with the procedure itself and the subsequent recovery. The estimates suggest that the disutility of catheterization is greater for those with worse Killip class. The estimates of parameters associated with the surgery choice are similar to those for the catheterization choice. The outside option is no transfer/no surgery. The intercept for no transfer/surgery is positive but relatively small in magnitude. The intercept for transfer/no surgery is negative and significant, indicating there is disutility associated with transfer. Transfer/surgery also has significant disutility. The disutility of surgery are greater for those with severe reductions in systolic functioning. It should be noted however that there are substantial benefits of surgery in terms of reduced mortality (Table 2). It is this expected benefit of surgery that provides an incentive for individuals to undergo catheterization and surgery despite the disutility associated with these procedures, and to choose a surgery hospital; a choice that also avoids the disutility of transfer.

2.10 Simulations

We use simulations from the model to generate a sample for which we know the value of treatment effect to compare the estimators described in section 3. In this section we describe the procedure used to simulate the data and compute the treatment effect based of the parameters of the estimated model.

The choice set is unique to each individual (see Section 2.7) implying that the dynamic programming problem has to be solved separately for each individual when simulating data. Given this computational burden, we randomly draw 1000 people once from the CCP data and using their associated period 1 variables (e.g., age, race, hospital choice sets) we draw error terms for each choice (catheterization, surgery) and each health transition (blockage, mortality). In this way we simulate choices and transitions for this sample of 1000 individuals. We repeat this 100 times for each person ending up with 100,000 simulated histories. In doing these simulations we assume that all individuals are type 1. This simulated sample is used to compare the estimators.

As seen in Table 1 (column labeled "Simulated Data"), the simulations do a very well in replicating the basic characteristics of the CCP data. The frequencies of outcomes (i.e., mortality, demographics, patient characteristics) and choices (i.e., catheterization, hospital service and volume categories) match the means in the data. Table 5 reports the characteristics for the simulated sample conditional on whether the simulated individual received catheterization or not.

Table 5 about here

We calculate average treatment effect (ATE) and average treatment effect on the treated (ATT) based on the structural model in the following way. ATE is calculated by simulating the history of each of the 1000 individuals under two different scenarios: once forcing *everyone to get* catheterization (irrespective of their initial hospital choice and state variables), and then again forcing *everyone to not get* catheterization. All other choices and transitions are made optimally given the model. The ATE is calculated as the difference in mean mortality in the sample between the two simulations:

$$ATE = \overline{m}_{cath} - \overline{m}_{nocath} \ ,$$

where $\overline{m}_{cath}$ is mean mortality forcing everyone to have catheterization and $\overline{m}_{nocath}$ is mean mortality forcing everyone not to have catheterization. ATT is computed in the same way, but the means are for the sample that chose to get catheterization when the catheterization choices were unconstrained.

The reported ATE and ATT are the mean ATE and ATT over 100,000 simulations. The "true" ATE and ATT are computed to be, -0.099 and -0.090 respectively.

3 Treatment Effects Estimators and the Structural Model

Unlike the structural discrete choice model of Section 2 which explicitly models all the utility maximizing decisions, the literature on treatment effects tries to estimate the causal effect of a single program or treatment choice on an outcome of interest. An implicit assumption is that by concentrating on a single parameter this causal effect of the treatment of interest can be identified under weaker conditions than in a structural model (Heckman and Vytlacil 2007a). In this Section we analyze under which conditions our structural model can identify and estimate the average treatment effect (ATE), and the average treatment effect on the treated (ATT) using matching estimators.

Assume a researcher wishes to estimate the causal effect of a patient receiving catheterization (d) on mortality (m) (see e.g., McClellan et al. 1994, McClellan and Newhouse 1997). The treatment literature assumes two hypothetical mortality outcomes depending on the choice of catheterization

$$\begin{aligned} m_1 &= \mu_1(Z, u_1) \text{ with catheterization } , \\ m_0 &= \mu_0(Z, u_0) \text{ without catheterization } , \\ d &= 1(Z, u_d) \ , \end{aligned}$$

where Z is a vector of covariates observed to the researcher, but u_1, u_0 and u_d are unobserved terms that affect outcomes and choices respectively. The average treatment effect (ATE) is the mean effect of catheterization on mortality for a randomly drawn person of the population, e,g.

$$ATE \equiv E_Z \left[E\left(m_1 - m_0 | Z \right) \right] \ . \tag{3.1}$$

The average treatment effect on the treated (ATT) is the mean effect of catheterization on mortality for those who chose catheterization:

$$ATT \equiv E_Z \left[E\left(m_1 - m_0 | Z, d = 1 \right) \right] \ . \tag{3.2}$$

Thus, ATT measures the expected difference in mortality from choosing catheterization for those who selected catheterization.

Various estimation strategies can be used to estimate the ATE and the ATT under different sets of assumptions. A full description and analysis of the performance of all methods and strategies used to estimate treatment effects is beyond the scope of this paper. Instead, we concentrate on estimators in two broad classes: matching estimators and instrumental variable estimators.[9] In the spirit of Heckman et al (2007b), we classify any estimator based on the conditional independence assumption (defined below) as a matching estimator. Other authors (Imbens 2004, Cameron and Trivedi 2005, and Wooldridge 2001) have classified treatment estimators differently.

3.1 Matching Estimators

Matching Estimators are based on two assumptions. The first assumption is conditional independence (also called unconfoundedness, ignorability of treatment or selection on observables in the matching literature),

$$(m_1, m_0) \perp d|Z \ ,$$

where $\perp$ denotes "independence" and implies that the catheterization decision (d) is independent of the two mortality outcomes (m_1, m_0) once we control for Z. This assumption can be relaxed to mean independence, which using our notation will be satisfied if,

$$(u_1, u_0) \perp u_d|Z \ .$$

In simple words, mortality may be correlated with the catheterization decision. However, once we partial out the effect of the controls Z, mortality and catheterization are uncorrelated. The second assumption is the overlap assumption,

$$0 < \Pr(d = 1|Z) < 1.$$

This assumption implies that in large samples, for each value Z, there are observations for which we observe m_1 and some other observations for which we observe m_0.[10]

[9] A third alternative is the method of controls. See Heckman and Vytlacil (2007b) for a description of this method and Aakvik el al (2005) for a method of control estimator when the dependent variable is discrete. We did not explore this method, because in our case it would be computationally prohibitive.

[10] If one is only interested in the ATT the one only needs to assume that $m_0 \perp d|Z$ and $\Pr(d = 1|Z) < 1$ (see e.g., Heckman, Ichimura, & Todd 1997 and Imbens 2004).

Matching estimators that we study are OLS, flexible OLS, flexible logits, and nearest neighbour propensity score matching.[11] All of these methods are simple to implement and widely used. In the Appendix we discuss the implementation of these methods. The main difference among these methods is how they impute the missing outcome. For those patients that undergo catheterization we need to impute the potential outcome without catheterization, and vice-versa. The first three methods use a regression function to impute the missing outcomes, while the fourth uses the outcome of the nearest neighbors of the opposite treatment group to impute the missing outcomes.[12]

From our structural model the two hypothetical potential outcomes variables ($j = 1$ catheterization and $j = 0$ no catheterization) are generated by:

$$m_j = 1\left(\gamma_0 + dd_s\gamma_1 + b\gamma_2 + k\gamma_3 + h\gamma_4 + D\gamma_5 + Xdd_s\gamma_6 + \epsilon_4\right) \tag{3.3}$$

where ϵ_4 follows a logistic distribution, d (no transfer/catheterization or transfer/catheterization) is the catheterization dummy, and d_s (no transfer/surgery or transfer/surgery) is the surgery dummy. In our model, after catheterization reveals the level of blockage, b, the decision maker has the option of selecting surgery. Thus, catheterization does not affect mortality directly but only indirectly as it affects the option for surgery later. Because γ_1 and γ_6 are negative (see Table 2), catheterization will never increase the probability of mortality. Of course many decision makers will not choose catheterization because the disutility associated with the procedure (see Table 4) outweighs the utility benefits of lower mortality.

In our model the decision maker chooses catheterization in period 2, and at that time, b and ϵ_4 are unknown and d_s is the period 3 treatment (surgery) decision. However, $I_2 = (d_1, h, k(t), D, X)$ are observed by the decision maker at the time of the catheterization decision. From equation (2.9), the

[11] A fully saturated model with all possible interactions will lead to a matching estimator providing a measure of the difference in the means between treated and untreated observation for each cell. Unfortunately, this method is unfeasible due to large number of cells. Instead we interact every explanatory variable with treatment, and some variables are also interacted among themselves as well as with the treatment variable. See Table 1 for the interactions in the demographic and other variables.

[12] Methods that use a regression function to extrapolate the missing outcomes do not need the overlap assumption. However, if Z perfectly predicts d, the treatment effects are not identified for those values of Z because there is no variation in d that is fully explained by Z.

optimal catheterization decision is given by:

$$\delta_2(I_2,\epsilon_2) = \arg\max_{d_2\in D_2(d_1)} \left\{\overline{V}_2(I_2,d_2)+\epsilon_2\right\},$$
$$\overline{V}_2(I_2,d_2) = \overline{U}_2(I_2,d_2)+\mu\sum_{b=1}^{3}\int_{\epsilon_3} V_3(b,\epsilon_3|d_2)\Pr(b|k,h,D)\,f(\epsilon_3)$$

where ϵ_2 represents unobserved heterogeneity, i.e., factors affecting the catheterization decision observed by the decision maker but not the researcher.

$$d=\begin{cases} 1 \quad if \quad \delta_2(I_2,\epsilon_2)=1 \text{ or } \delta_2(I_2,\epsilon_2)=3 \\ 0 \qquad\qquad otherwise \end{cases}.$$

The utility function $(\overline{V}_2(I_2,d_2)+\epsilon_2)$ follows the additively separately decomposition with respect to ϵ_2, (assumption AS of Rust (1994)). Also since the unobserved heterogeneity ϵ_2 is uncorrelated with ϵ_3, ϵ_4, and b, the structural model also satisfies the conditional independence assumption (Assumption CI of Rust (1994)).

Assuming the researcher observes I_2, the only source of unmeasured heterogeneity to the researcher is ϵ_2. Although ϵ_2 influences the catheterization decision it will be uncorrelated with future realized values of b, and ϵ_4. Furthermore, it will only affect d_3 through the choice of d_2 and the conditional independence assumption holds. Finally, the overlap assumption is always satisfied because ϵ_2 has a continuous distribution with unbounded support. Thus any matching estimator based on the assumptions of conditional independence and overlap should be able to identify the ATE and ATT using I_2 as the conditioning set, since the AS-CI assumptions of the structural model, and the unboundedness in the support of the unmeasured heterogeneity ensure that these assumptions are satisfied. In other words, the AS-CI assumptions imply that conditioning on I_2, ϵ_2 serves as randomization mechanism of d with respect to (m_1,m_0). For a given value of I_2, a particular draw of ϵ_2 may lead a decision maker to choose $d=1$, while a different draw may lead him to choose $d=0$.[13]

Moreover, under the assumptions of conditional independence and overlap the average treatment is the same as the marginal treatment which leads to $ATE(I_2)=ATT(I_2)$ as shown in Heckman et al (1997) and Heckman and Vytlacil (2007). However, the unconditional treatment effects are different,

[13]In fact, it can also be shown that the AS-CI assumptions are satisfied for the utility functions and the transition probabilities in all the time periods of our structural model, and that any ATE and ATT based on decisions at time j with conditioning set I_j can be consistently estimated using matching estimators.

$ATE \neq ATT$.

In practice it is likely that factors which are correlated with future values of the state variables are also missing. For example, most researchers who study the effect of catheterization do not observe detailed measures of severity like the Killip class, k. By definition k is correlated with b and it will also affect mortality directly, which implies that the conditional independence assumption is violated. By a similar argument it can be shown that the conditional independence assumption of a treatment estimator will not be satisfied if d_1, h, D, and X are not observed. Thus, making the CI assumption about the unobserved heterogeneity term in the structural model is no more restrictive than the conditional independence assumption made when using matching estimators.[14]

3.2 Instrumental Variables

When the conditional independence assumption of the matching estimators fails instrumental variable methods are commonly used. The matching estimators do not distinguish between variables that affect the outcomes and variables that only affect the treatment equation. The existence of a variable that directly affects the treatment decision but not the outcomes directly is the key concept underlying instrumental variable methods.

In our structural model *dist* affects treatment (see equation 2.5), but does not directly affect mortality satisfying the traditional conditions for a good instrument.[15] Unfortunately, in our simulated data the response to treatment is heterogeneous and in this case standard instrumental variable methods do not estimate ATE or ATT. If the instruments satisfy the additional assumption of monotonicity (Imbens and Angrist 1994), then instrumental variable methods will estimate the local average treatment effect ($LATE$), which is the mean gain to those who would be induced to undergo catheterization because of a change in the instrumental variable. However, our instruments do not satisfy the monotonicity assumption that a change in the instrumental variable will have a monotonic effect on the treatment decision. As discussed in Heckman and Vytlacil (2005) when the treatment decision is also heterogenous, the method of instrumental variables does not lead to economically interpretable treatment effects. Despite this, we apply linear two-stage least squares (TSLS) with a common treatment effect since these methods are widely used in the empirical literature. We employ a linear TSLS where

[14] We appreciate the comments of the editor John Rust in writing this section.

[15] In our simulated data we fixed the unobserved patient type. Thus, d_1 is also a valid intrument. However, we used *dist* as the instruments because this instrument is widely applied in the literature. See McClellan and Newhouse 1997.

treatment is allowed to vary with Z, and a bivariate probit to our simulated data using distance to different type of hospitals as an instrument.

4 Empirical Specification for Treatment Effect Regressions

In this section we describe the regression specifications we employ for comparing the estimators described in section 3 using the simulated data from the structural model.

4.1 Dependent Variables

For all of our specifications the dependent variable is a binary indicator for mortality at one year following AMI. This is the least controversial measure of quality of care used in the health economics literature.

4.2 Treatment Variable

Our treatment variable is cardiac catheterization, which has been widely studied in health economics. We define a binary variable indicating whether the patient received catheterization. We adopt a linear specification because it is common in the literature but allow for some interactions as described below.

4.3 Controls

We divide our controls into different groups. The first group only contains a dummy for treatment. The second group includes demographic characteristics, which is information contained in most studies. The third group adds hospital characteristics, which is information that is found with typical hospital administrative data. The fourth group adds controls for severity of illness which is information that is found on patient charts and is typically not available.

Demographic characteristics: Demographic variables include age at the time of the admission, gender, and race. We classify age into three categories 65-74, 75-84, and 85+. We include MSA[16] fixed effects as it has been done in the literature. We also *interact* the age, gender, and race variables to create a fully interacted model.

[16] A metropolitan statistical area (MSA) is a core area containing at least one urbanized area of 50,000 or more inhabitants, together with adjacent communities having a high degree of economic and social integration with that core.

Hospital characteristics: We use the following controls for hospital characteristics. An indicator variable, "Service," designates one of three possibilities for each hospital and is coded as: "1 = neither catheterization nor surgery services available (omitted category)," "2= catheterization available but not surgery," "3 = both catheterization and surgery available." A second indicator variable accounts for the annual "Volume" of surgeries and takes on one of four values: "1 = 0-49 surgeries (omitted category)," "2 = 50 to 99 surgeries," "3 = 100 to 199 surgeries,"and "4 = 200+ surgeries."

Severity of illness and coexisting conditions: An advantage of CCP compared to Medicare claims data and standard surveys is that it contains very detailed information on patients' coexisting conditions (co-morbidities) and severity of illness. In our regressions we use indicator variables for the Charlson index being in one of three categories (category one is omitted category). The regressions include indicator variables for the Killip index being in one of three categories, i.e., (i) class 1 (omitted category), or (ii) class 2, or (iii) class 3 or 4.

Instruments: Our main instrumental variable for catheterization is the distance between the closest catheterization hospital to the patient's zip-code centroid. The justification is that, everything else equal, a patient will select the closer hospital, and the farther is a catheterization hospital the less likely the patient will be admitted to a catheterization hospital. We use the same distance classifications in creating our IV as in the structural model (i.e., 10 categories defined by distance to nearest hospital, see Table 3 for the specific distance categories). McClellan et al. (1994) originally proposed differential distance between the closest catheterization hospital and the closest non-catherterization hospital from the patients zip-code as an instrument for catheterization. They also analyzed in detailed the properties of differential distance as an instrument for catheterization. McClellan and Noguchi (2001) analyzed whether differential distance is correlated with the detailed severity of illness measures contained in CCP data and concluded that it is not correlated. Our IV is slightly different from the one analyzed by McClellan et al. (1994) because it has stronger predictive power than differential distance given the specification of our structural model. For all our estimators standard errors were bootstrapped with 500 repetitions for samples of 1000 observations each, with one draw for each person drawn out of the 100,000 observations in the total simulated sample. We note that one needs to exercise caution in interpreting the bootstrapped standard errors for the nearest neighbor matching estimator since, as shown in Abadie and Imbens (2009) for the continuous explanatory variable case, these standard errors may be misleading.

5 Results: Comparison of Estimators

In evaluating estimators described in section 3, we compare the effects of catheterization. This is done using four different sets of regressors to account for different kinds of data that may be available to the researcher: (i) the first specification includes only an indicator for treatment (Table 6, Col. 1), (ii) a second specification includes in addition demographic information about individuals (Col. 2), (iii) a third specification further includes measures of hospital characteristics (Col. 3), and (iv) the fourth and most comprehensive specification also includes detailed measures of severity of illness (which can only be found on hospital charts) along with the other regressors (Col. 4).

Table 6 about here

Table 6, Panel A presents the results for estimating the ATE. For matching methods, it is seen that the estimates for the first three specifications (Cols. 1-3) are broadly similar. The ATE is however estimated incorrectly (recall that the "true" ATE is -0.099). In the fourth specification (Col. 4), the performance is much improved. The Flexible Logit estimator is closest to the true ATE, with the Fully Interacted OLS, Nearest Neighbor Propensity Score Matching and OLS Matching Estimators following in that order. Methods using instruments show a lot of variation in performance in recovering the true ATE. For the 2SLS, the performance improves as measures of heterogeneity are added. However, specifications 3 and 4 overestimate and underestimate the true ATE by approximately the same amount. The fully interacted 2SLS performs the worst among all the estimators in Panel A. The results of the Bivariate Probit improve from specification 1 to 3, but get worse when the richest specification (Col. 4) is used. In fact, it exactly estimates the true ATE in specification 3.

Table 6, Panel B repeats the exercise performed in Panel A, but now the object of interest is ATT. The estimates of ATT exhibit the same patterns as seen for the estimates of ATE in Panel A, with one exception. The results for the Bivariate Probit are poor in contrast with Panel A. A reason for the similarity of the ATT results to those for ATE could be that the distribution of treated people is similar to the full sample. If both of these are random samples then one would expect the estimates to be the same theoretically. The more dissimilar the treated group from the full group, i.e., the greater the selection problem, the larger the expected difference between the two estimates. Hence either (a) the treated sample is a large part of the overall sample, (b) the samples are similar in observables, or (c) the treatment effect varies only slightly with observables.

In sum, we find that the estimators do not perform well in recovering the "true" effect (be it ATE or ATT) when the data are poor in measures of individual heterogeneity, i.e., detailed patient characteristics in our case (specification 4). However, when the data are rich in measures of such characteristics matching methods perform well. Not surprisingly, when detailed patient characteristics are missing methods using instruments do a poor job in recovering the "true" ATE or ATT. However, as we discussed in Section 3.2, instrumental variable methods at best recover LATE in this scenario. For the same reason, when detailed patient characteristics are included instrumental variable methods perform better, but not as well as matching methods.

6 Conclusions

In this study we compare different estimators widely used in the treatment effects literature with respect to their ability to estimate the effects of catheterization on survival outcomes for individuals with AMI. In the *absence* of a measure of the *true* treatment effect we adopt a novel strategy to compare the estimators. We estimate a structural model of treatment and hospital choices for individuals with acute myocardial infarction (AMI) using a unique data set that combines administrative Medicare claims data with hospital chart data on patient severity of illness. With the estimated structural model we simulate data for which the treatment effect is known to us, and use this to evaluate the performance of the treatment effects estimators. We focus on examining the performance of matching estimators in recovering ATE and ATT. We show that in the context of our structural model when the researcher has information on detailed patient characteristics, the assumptions underlying matching estimators are satisfied. However, if either detailed patient characteristics or hospital characteristics are not observed then the assumptions underlying treatment estimators are not satisfied. Matching estimators perform well when measures of heterogeneity like patient characteristics are included in the regression specification. However, when the data is poor in such measures of individual heterogeneity then the estimators do a poor job in recovering the true treatment effect.

Another novelty of our analysis is that we use Medicare claims data combined with data from the Cooperative Cardiovascular Project (CCP) merged with the American Hospital Association's (AHA) annual survey of hospitals. One significant limitation of the Medicare claims data used previously to study these treatment effects is that it does not contain comprehensive measures of severity of illness. Thus, another contribution of our study is to use a richer data set to analyze how sensitive different estimators are to the addition of detailed severity of illness measures. Severity of illness is a primary factor driving selection into treatments, and if omitted leads to bias in the estimation of the treatment-outcome relationship. We find that the estimators do a poor job in recovering the true treatment effect when the data are poor in measures of individual heterogeneity. As measures of heterogeneity are added to the regression specification the results improve, especially for matching estimators.

There are two important implications of this analysis for researchers using hospital Medicare claims data or other types of hospital claims data. First, we find that distance does not perform well as an instrument in estimating treatment effects (ATE and ATT) in our application. These types of results have been seen before in the literature on weak instruments (e.g., Bound, Jaeger, and Baker 1995; Staiger and Stock 1997). Second, we do not find evidence that estimation methods assuming selection on observables perform well in estimating treatment effects in the *absence* of good data. Hence, researchers should be careful when estimating causal effects using hospital claims data. The best situation is to have detailed data on severity of illness and coexisting conditions for at least a sub-set of the sample. Distance between the patients' residence and different types of hospitals, which has been considered an excellent instrument for the kind of applications examined in this paper, should be used with caution. Finally, further investigation of the performance of treatment effect estimators in other settings is warranted.

References

[1] Aakvik A, Heckman J, Vytlacil E. 2005. Estimating treatment effects for discrete outcomes when response to treatment vary: an application to Norwegian vocational rehabilitation programs. *Journal of Econometrics* 125: 15-51.

[2] Abadie A, Imbens G. 2009. On the failure of bootstrap for matching estimators. *Econometrica* 76: 1537-1557.

[3] Angrist JD. 2001. Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *Journal of Business and Economic Statistics* 19: 2-16.

[4] Angrist JD, Krueger AB. 1999. Empirical strategies in labor economics. In *Handbook of Labor Economics Vol 3*, Ashenfelter O, Card D (eds). Elsevier: Amersterdam.

[5] Arcidiacono P, Jones JB. 2003. Finite mixture distributions, sequential likelihood and the EM algorithm. *Econometrica* 71: 933-946.

[6] Arcidiacono P, Khwaja A, Ouyang L. 2007. Habit persistence and teen sex: could increased access to contraception have unintended consequences for teen pregnancies? Mimeo, Department of Economics, Duke University.

[7] Barnow B, Cain G, Goldberg A. 1980. Issues in the analysis of selectivity bias. *Evaluation Studies* 5: 42-59.

[8] Bound J, Jaeger D, Baker R. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90: 443-450.

[9] Cameron AC, Trivedi PK. 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press: New York, NY.

[10] Charlson ME, Pompei P, et al. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases* 40: 373-383.

[11] Cobb-Clark D, Crossley T. 2003. Econometrics for evaluation: an introduction to recent developments. *Economic Record* 79: 491-511.

[12] Dehejia R, Wahba S. 1999. Causal effects in nonexperimental studies: reevaluating the evaluation training programs. *Journal of the American Statistical Association* 94: 1053-1062.

[13] Gowrisankaran G, Town R. 1999. Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics* 18: 747-767.

[14] Hamilton B, Hamilton VH. 1998. Estimating surgical volume-outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Economics* 6: 383-95.

[15] Heckman J, LaLonde R, Smith J. 1999. The economics and econometrics of active labor market programs. In *Handbook of Labor Economics Vol 3*, Ashenfelter O, Card D (eds). Elsevier: Amsterdam.

[16] Heckman J, Ichimura H, Todd P. 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training program. *Review of Economic Studies* 64: 605-654.

[17] Heckman J, Ichimura H, Todd P. 1998. Matching as an econometric evaluation estimator. *Review of Economic Studies* 65: 261-294.

[18] Heckman J, Vytlacil E. 2005. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73: 669-738.

[19] Heckman J, Vytlacil E. 2007a. Econometric evaluation of social programs, part I: causal models, structural models, and econometric policy evaluation. In *Handbook of Econometrics Vol 6.B*, Heckman J, Leamer E (eds). Elsevier: Amsterdam.

[20] Heckman J, Vytlacil E. 2007b. Econometric evaluation of social programs, part II: using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In *Handbook of Econometrics Vol 6.B*, Heckman J, Leamer E (eds). Elsevier: Amsterdam.

[21] Ho V. 2002. Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. *Journal of Health Economics* 21: 873-885.

[22] Imbens G. 2004. Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics* 86(1): 4-29.

[23] Imbens G, Angrist JD. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62(2): 467-475.

[24] Khwaja A. 2001. Health insurance, habits and health outcomes: a dynamic stochastic model of investment in health. Unpublished Ph.D. dissertation, University of Minnesota.

[25] Killip T, Kimbal JT. 1967. Treatment of myocardial infarction in a coronary care unit. *American Journal of Cardiology* 20: 457-464.

[26] Krumholz HM, Chen J, Chen YT, et al. 2001. Predicting one-year mortality among elderly survivors of hospitalization for an acute myocardial infarction: results from the Cooperative Cardiovascular Project. *Journal of the American college of Cardiology* 38: 453-9.

[27] LaLonde RJ. 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76: 604-620.

[28] Luft HS, Garnick D, Mark D, McPhea S. 1990. *Hospital Volume; Physician Volume and Patient Outcomes: Assessing Evidence*. Health Administration Press Perspectives: Ann Arbor, MI.

[29] Marciniak TA, Ellerbeck EF, Radford MJ, et al. 1998. Improving the quality of care for Medicare patients with acute myocardial infarction: results from the Cooperative Cardiovascular Project. *JAMA* 287: 1269-1276.

[30] McClellan M, Newhouse JP. 1997. The marginal cost-effectiveness of medical technology: a panel instrumental-variables approach. *Journal of Econometrics* 77: 39-64.

[31] McClellan M, McNeil B, Newhouse JP. 1994. Does more intensive treatment of acute myocardial infarction reduce mortality? *JAMA* 272: 859-866.

[32] McClellan M, Noguchi H. 2001. Validity and interpretation of treatment effect estimates: using observational data. Stanford University, Unpublished Manuscript.

[33] McFadden D. 1981. Econometric models of probabilistic choice. In *Structural Analysis of Discrete Data with Econometric Applications*,Manski C, McFadden D (eds). MIT Press: Cambridge, MA.

[34] Michalopoulos C, Bloom HS, Hill CJ. 2002. Can propensity-score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? *Review of Economics and Statistics* 86(1): 156-179.

[35] Rosenbaum P, Rubin D. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41-55.

[36] Rubin D. 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & sons: New York.

[37] Rust J. 1994. Structural estimation of markov decision processes. In *Handbook of Econometrics, Vol 4*, Engle RF, McFadden DL (eds). Elsevier: Amsterdam.

[38] Rust J, Phelan C. 1997. How Social Security and Medicare affect retirement behavior in a world of incomplete markets. *Econometrica* 65: 781-831.

[39] Sloan FA, Picone G, Taylor D, Chou S. 2001. Hospital ownership and cost quality of care: is there a dime's worth of difference? *Journal of Health Economics* 20: 1-21.

[40] Staiger D, Stock J. 1997. Instrumental variables estimation with weak instruments. *Econometrica* 65: 557-586.

[41] Trogdon JG. 2009. Demand for and regulation of cardiac services. *International Economic Review*, in press.

[42] Wooldridge J. 2001. *Econometrics Analysis of Cross Section and Panel Data.* MIT Press: Cambridge, MA.

7 Appendix

7.1 Matching Estimators

1. Ordinary Least Squares (OLS) and Flexible OLS: Assume that the outcome equations are given by a linear probability model

$$\begin{aligned} m_0 &= Z'\beta + \epsilon \qquad if \quad d = 0 \\ m_1 &= Z'\beta + \delta + \epsilon \quad if \quad d = 1 \end{aligned} \tag{7.1}$$

then under the conditional independence assumption, we can apply least squares to

$$m = Z'\beta + d\delta + \epsilon \tag{7.2}$$

where $\widehat{ATE}$=$\widehat{ATT}$=$\widehat{\delta}$. We also allow for heterogenous effects by estimating a flexible linear probability model

$$m = Z'\beta_0 + d\left(Z - \overline{Z}\right)'\gamma + d\delta + \epsilon \tag{7.3}$$

where$\widehat{ATE} = \widehat{\delta}$ and $\widehat{ATT} = \widehat{\delta} + \left(\sum_{i=1}^{N} d_i\right)^{-1} \sum_{i=1}^{N} d_i \left(Z_i - \overline{Z}\right)\widehat{\gamma}$. See Barnow, Cain, and Golberger (1980) and Wooldridge (2001). All the components of Z are discrete and Z includes the main interaction effects among the demographic variables. Although this model is not fully saturated, it is very general and should recover the ATE and ATT if the ignorability of treatment assumption is satisfied.

2. Flexible Logit: Split the sample between individuals with treatment ($d = 1$) and without treatment ($d = 0$). A natural extention of the linear probability model given above is to assume that $m_0 = 1\left(Z'\beta_0 + \epsilon_0\right)$ and $m_1 = 1\left(Z'\beta_1 + \epsilon_1\right)$ where ϵ_0 and ϵ_1 are independent logistic distributions, and the estimate separate logits for individuals with and without treatment. Then, simple estimators of the ATE and ATT are given by

$$\widehat{ATE} = N^{-1}\sum_{i=1}^{N}\left[\Lambda\left(Z_i'\widehat{\beta}_1\right)-\Lambda\left(Z_i'\widehat{\beta}_0\right)\right]$$

$$\widehat{ATT} = \left(\sum_{i=1}^{N}d_i\right)^{-1}\sum_{i=1}^{N}d_i\left[\Lambda\left(Z_i'\widehat{\beta}_1\right)-\Lambda\left(Z_i'\widehat{\beta}_0\right)\right]$$

where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are logits estimates of the covariates for individual with and without catheterization and $\Lambda(.)$ is the logistic cdf. In other words, for those individuals who received treatment we use $\Lambda\left(Z_i'\widehat{\beta}_0\right)$ to impute the missing outcome and for those individuals who did not receive treatment we use $\Lambda\left(Z_i'\widehat{\beta}_1\right)$ to impute the missing outcome.

3. Nearest Neighbor Propensity Score Matching: We estimate the nearest neighbor propensity score matching developed by Rosenbaum and Rubin (1983). Let i be an element of the treatment group, $i \in \{d=1\}$, j be an element of the comparison group, $j \in \{d=0\}$, and $p(Z_i) = \Pr(d=1|Z_i)$ the propensity score. We implement the method as follows:

1. (a) For each $i \in \{d=1\}$ find the closest neighbor j from the control group. The closest neighbor is defined as the closest propensity score:

$$\left\{j| \min_{j\in\{d=0\}} |p(Z_i)-p(Z_j)|\right\}$$

(b) Calculate the matched outcome as the value of the outcome of the nearest neighbor in the comparison group:

$$\widehat{E}(m_{0i}|p(Z_i), d_i=0) = m_{0j}$$

(c) Calculate the

$$\widehat{ATT} = \frac{1}{N_1}\sum_{i\in\{d=1\}}\left[m_{1i}-\widehat{E}(m_{0i}|p(Z_i), d_i=0)\right]$$

(d) To calculate the ATE we also need to $\widehat{E}(m_{1j}|p(Z_i), d_i=1)$ for those individuals in the comparison group.

7.2 Instrumental Variable Methods

4. Linear Two Stage Least Squares (TSLS): We apply TSLS to the linear probability model described in equation (7.2) using the propensity score as the instrument and the hospital distance dummies as the exclusion restrictions. We also apply TSLS to the linear probability model given in equation (7.3) using $\Pr(d = 1|Z, dist) \times Z$ at the additional instruments.

5. Bivariate Probit: Finally, we also estimate the model using a bivariate probit and report marginal effects for catheterization. This model assumes that

$$\begin{aligned} m &= Z'\beta_1 + d\delta + \epsilon_1 \\ d &= Z'\beta_2 + dist \times \gamma + \epsilon_2 \end{aligned}$$

where (ϵ_1, ϵ_2) follow a bivariate normal distribution. We estimate this model using the standard full information maximum likelihood estimator for this problem.

Table 1: Sample Statistics

	CCP data		Simulated Data	
	Mean	Std. Dev	Mean	Std. Dev
Outcome variable				
1-year mortality	0.336	(0.472)	0.325	(0.468)
Treatment variable				
Catheterization 90 days	0.348	(0.476)	0.353	(0.477)
Demographic characteristics				
Age 65 to 74	0.222	(0.416)	0.225	(0.417)
Age 75 to 84	0.181	(0.385)	0.173	(0.378)
Age 85 and over	0.056	(0.229)	0.064	(0.245)
Female age 65 to 74	0.150	(0.357)	0.151	(0.358)
Female age 75 to 84	0.194	(0.395)	0.200	(0.400)
Female age 85 and over	0.097	(0.295)	0.091	(0.288)
Black age 65 to 74	0.025	(0.156)	0.020	(0.140)
Black age 75 to 84	0.015	(0.123)	0.014	(0.117)
Black age 85 and over	0.005	(0.071)	0.004	(0.063)
Female black age 65 to 74	0.025	(0.155)	0.030	(0.171)
Female black age 75 to 84	0.021	(0.143)	0.020	(0.140)
Female black age 85 and over	0.009	(0.093)	0.008	(0.089)
Non metropolitan area	0.259	(0.438)	0.259	(0.438)
Metropolitan area	0.741	(0.438)	0.741	(0.438)
Hospital characteristics				
No service hospital	0.343	(0.475)	0.360	(0.480)
Catheterization hospital	0.227	(0.419)	0.217	(0.412)
Surgery hospital	0.430	(0.495)	0.423	(0.494)
Surgery volume 0 to 49	0.653	(0.476)	0.659	(0.474)
Surgery volume 50 to 99	0.157	(0.364)	0.161	(0.367)
Surgery volume 100 to 199	0.135	(0.341)	0.121	(0.326)
Surgery volume 200 and over	0.055	(0.228)	0.059	(0.235)
Detailed patient characteristics				
Charlson index 1	0.508	(0.500)	0.524	(0.499)
Charlson index 2 – 3	0.356	(0.479)	0.354	(0.478)
Charlson index > 3	0.136	(0.343)	0.122	(0.327)
Killip Class 1	0.494	(0.500)	0.526	(0.499)
Killip class 2	0.119	(0.324)	0.109	(0.311)
Killip class 3	0.387	(0.487)	0.365	(0.418)
No Blockage	0.045	(0.207)		
Blockage moderate	0.171	(0.376)		
Blockage severe	0.023	(0.150)		
Blockage missing	0.761	(0.426)		
N	114,818		100,000	

Table 2: Health Transitions: Multinomial Logits[1]

	Killip Class			Blockage			Mortality		
Variable	Coeff.	SE	OR	Coeff.	SE	OR	Coeff.	SE	OR
	Killip II			Moderate Reduction			Dead at 1 year		
Intercept	-1.845*	0.022		1.163*	0.028		-1.906*	0.051	
Age 75 to 84	0.298*	0.022	1.35	0.075*	0.034	1.08	0.453*	0.016	1.57
Age 85 and over	0.566*	0.030	1.76	0.165	0.095	1.18	0.916*	0.022	2.50
Female	0.078*	0.019	1.08	-0.265*	0.032	0.77	-0.014	0.015	0.99
Minority	0.092*	0.032	1.10	0.015	0.054	1.02	-0.044	0.023	0.96
Charlson 2-3	0.292*	0.021	1.34	0.263*	0.036	1.30	0.288*	0.016	1.33
Charlson > 3	0.588*	0.030	1.80	0.272*	0.064	1.31	0.616*	0.022	1.85
Killip II				0.216*	0.050	1.24	0.313*	0.023	1.37
Killip III or IV				0.654*	0.045	1.92	0.779*	0.017	2.18
Moderate reduction							0.423*	0.059	1.53
Severe reduction							1.249*	0.069	3.49
Surgery							-1.124*	0.029	0.32
Surgery*volume 50-99							-0.118*	0.030	0.89
Surgery*volume 100-199							-0.043	0.029	0.96
Surgery*volume 200+							-0.267*	0.050	0.77
Type 2 intercept	0.066	0.069							
	Killip III or IV			Severe Reduction					
Intercept	-1.096*	0.015		-1.549*	0.052				
Age 75 to 84	0.460*	0.014	1.58	0.079	0.051	1.08			
Age 85 and over	0.874	0.020	2.40	0.365*	0.122	1.44			
Female	0.196*	0.013	1.22	-0.537*	0.049	0.58			
Black	0.184*	0.021	1.20	0.164*	0.077	1.18			
Charlson 2-3	0.740*	0.014	2.10	0.809*	0.054	2.25			
Charlson > 3	1.207*	0.020	3.34	1.049*	0.082	2.85			
Killip II				0.634*	0.081	1.89			
Killip III or IV				1.997*	0.059	7.37			
Type 2 intercept	-0.447*	0.056	1.05						
Probability of Type 2	0.094								
N					114,818				
ln L					-4.65E+05				
Likelihood ratio index					0.390				

Asymptotic standard errors (SE) and odds ratios (OR) are reported.

* indicates significance at the 95% confidence level.

[1] The reference group for each transition is the healthiest category: Killip I, normal systolic function, and alive at one year.

Table 3: Hospital Utility Parameters[1]

Variable	Coeff.	SE
Cath only hospital	-3.450*	0.276
Cath only*Charlson 2-3	0.452*	0.041
Cath only*Charlson > 3	0.759*	0.062
Surgery hospital	-2.429*	0.221
Surgery*Charlson 2-3	0.509*	0.040
Surgery*Charlson > 3	0.820*	0.059
Distance 4.7-8.5 mi.	-1.320*	0.047
Distance 8.6-12.9 mi.	-2.291*	0.040
Distance 13.0-18.7 mi.	-3.464*	0.038
Distance 18.8-25.1 mi.	-4.674*	0.040
Distance 25.2-33.2 mi.	-5.844*	0.046
Distance 33.3-45.4 mi.	-7.177*	0.056
Distance 45.5-61.7 mi.	-8.538*	0.070
Distance 61.8-85.6 mi.	-9.579*	0.083
Distance 85.7+ mi.	-11.280*	0.110
Dist 4.7-8.5 mi.*MSA	-0.512*	0.049
Dist 8.6-12.9 mi.*MSA	-1.059*	0.043
Dist 13.0-18.7 mi.*MSA	-1.290*	0.045
Dist 18.8-25.1 mi.*MSA	-1.472*	0.053
Dist 25.2-33.2 mi.*MSA	-1.585*	0.067
Dist 33.3-45.4 mi.*MSA	-1.567*	0.092
Dist 45.5-61.7 mi.*MSA	-1.093*	0.125
Dist 61.8-85.6 mi.*MSA	-0.786*	0.163
Dist 85.7+ mi.*MSA	0.620*	0.187
Type 2*Cath only	5.910*	0.184
Type 2*Surgery	7.395*	0.173
$m(r_2/r_1)$	11.247*	0.825
N	114,818	
ln L	-4.65E+05	

SE Asymptotic standard errors.

* indicates significance at the 95% confidence level.

[1] Utility parameters in the first period are relative to choosing a hospital with no specialized services.

Table 4: Catheterization and Surgery Utility Parameters[1]

Variable	Cath Coeff.	Cath SE	Surgery Coeff.	Surgery SE
	No Tran/Cath		No Tran/Surgery	
Intercept	-1.121*	0.082	0.180*	0.032
Killip II	-0.493*	0.026		
Killip III or IV	-1.077*	0.019		
Moderate reduction			-0.036	0.038
Severe reduction			-0.886*	0.054
	Tran/No Cath		Tran/No Surgery	
Intercept	-3.228*	0.037	-1.103*	0.075
Killip II	-0.124	0.085		
Killip III or IV	-0.044	0.052		
Moderate reduction			0.089	0.085
Severe reduction			-0.166	0.129
	Tran/Cath		Tran/Surgery	
Intercept	-2.039*	0.086	-0.848*	0.066
Killip II	-0.430*	0.034		
Killip III or IV	-0.934*	0.024		
Moderate reduction			0.004	0.076
Severe reduction			-0.944*	0.136
$m(r_3/r_2)$	1.466*	0.105		
N		114,818		
ln L		-4.65E+05		

SE Asymptotic standard errors.

* indicates significance at the 95% confidence level.

[1] Utility parameters in the second and third periods are relative to choosing no transfer/no procedure.

Table 5: Characteristics for Treatment Group and Control Group for Simulated Sample

	Treatment Group		Control Group	
	Mean	Std. Dev	Mean	Std. Dev
Outcome variable				
1-year mortality	0.233	(0.423)	0.376	(0.484)
Demographic characteristics				
Age 65 to 74	0.246	(0.430)	0.213	(0.409)
Age 75 to 84	0.171	(0.376)	0.174	(0.379)
Age 85 and over	0.059	(0.237)	0.066	(0.248)
Female age 65 to 74	0.150	(0.358)	0.151	(0.358)
Female age 75 to 84	0.197	(0.398)	0.201	(0.400)
Female age 85 and over	0.086	(0.281)	0.093	(0.290)
Black age 65 to 74	0.021	(0.144)	0.019	(0.137)
Black age 75 to 84	0.012	(0.112)	0.014	(0.120)
Black age 85 and over	0.002	(0.053)	0.004	(0.067)
Female black age 65 to 74	0.024	(0.155)	0.032	(0.178)
Female black age 75 to 84	0.017	(0.129)	0.021	(0.145)
Female black age 85 and over	0.008	(0.089)	0.007	(0.088)
Non metropolitan area	0.225	(0.418)	0.277	(0.447)
Metropolitan area	0.774	(0.418)	0.722	(0.447)
Hospital characteristics				
No service hospital	0.224	(0.417)	0.433	(0.495)
Catheterization hospital	0.269	(0.443)	0.188	(0.391)
Surgery hospital	0.505	(0.499)	0.377	(0.484)
Surgery volume 0 to 49	0.621	(0.485)	0.679	(0.466)
Surgery volume 50 to 99	0.175	(0.380)	0.152	(0.359)
Surgery volume 100 to 199	0.134	(0.341)	0.113	(0.317)
Surgery volume 200 and over	0.068	(0.252)	0.053	(0.225)
Detailed patient characteristics				
Charlson index 1	0.559	(0.496)	0.504	(0.499)
Charlson index 2 – 3	0.336	(0.472)	0.363	(0.481)
Charlson index > 3	0.104	(0.305)	0.131	(0.338)
Killip Class 1	0.663	(0.472)	0.450	(0.497)
Killip class 2	0.105	(0.306)	0.111	(0.314)
Killip class 3	0.231	(0.421)	0.437	(0.496)
N	35,301		64,699	

Table 6: Estimation Using Simulated Data

	Catheterization indicator only	Demographic characteristics	Hospital characteristics	Detailed patient characteristics
	(1)	(2)	(3)	(4)
Panel A: Average treatment effect				
Matching Methods				
OLS	-0.140	-0.136	-0.136	-0.088
	(0.028)	(0.028)	(0.029)	(0.029)
Flexible OLS	-0.140	-0.136	-0.137	-0.091
	(0.028)	(0.028)	(0.029)	(0.031)
Flexible Logit	-0.186	-0.169	-0.166	-0.101
	(0.036)	(0.036)	(0.036)	(0.040)
Nearest Neighhbor Propensity Score	-0.140	-0.137	-0.138	-0.112
	(0.028)	(0.029)	(0.032)	(0.032)
Methods Using Instruments				
2SLS	0.023	-0.044	-0.128	-0.071
	(0.214)	(0.227)	(0.274)	(0.244)
Flexible 2SLS	0.023	-0.180	0.318	0.266
	(0.214)	(1.792)	(11.005)	(7.670)
Bivariate Probit	0.074	-0.022	-0.099	-0.067
	(0.273)	(0.280)	(0.290)	(0.293)
Panel B: Average treatment effect on the treated				
Matching Methods				
Flexible OLS	-0.140	-0.135	-0.136	-0.086
	(0.028)	(0.029)	(0.030)	(0.030)
Flexible Logit	-0.186	-0.169	-0.164	-0.092
	(0.036)	(0.036)	(0.037)	(0.035)
Nearest Neighhbor Propensity Score	-0.140	-0.136	-0.135	-0.092
	(0.028)	(0.029)	(0.035)	(0.037)
Methods Using Instruments				
Flexible 2SLS	0.023	-0.178	0.420	0.191
	(0.214)	(1.782)	(13.039)	(6.377)
Bivariate Probit	0.116	0.035	-0.026	0.003
	(0.242)	(0.236)	(0.233)	(0.238)
Panel C: "True" effects (with all covariates)				
Average treatment effect (ATE)	-0.099			
Average treatment effect on the treated (ATT)	-0.090			

Bootstrapped standard errors are reported in parentheses. For the list of variables included in demographic characteristics, hospital characteristics, and detailed patient characteristics see Table 1. The instruments are the distances to the nearest hospital with catheterization capability in the same categories as in Table 3.