



# Causal Effect of Ambulatory Specialty Care on Mortality Following Myocardial Infarction: A Comparison of Propensity Score and Instrumental Variable Analyses

MARY BETH LANDRUM

*Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, USA*

JOHN Z. AYANIAN

*Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA and Department of Medicine, Division of General Medicine, Brigham and Women's Hospital and Harvard Medical School*

*Received December 12, 2000; revised August 2, 2001; accepted October 5, 2001.*

**Abstract.** The quality and outcomes of care provided by primary care physicians and specialists are increasingly important issues in health policy research. Estimating the effect of specialty care on patient outcomes however is complicated by the observational nature of the studies. Patients treated by specialists are often different in terms of observed and unobserved characteristics that can bias estimates of specialty effects. We illustrate and compare two different analytic approaches, propensity scores and instrumental variables, to infer the causal effect of cardiology care in the ambulatory setting on 18-month mortality among 5467 elderly patients who survived at least 3 months after being hospitalized for a myocardial infarction in New York state during 1994 and 1995. Using both approaches we found reductions in 18-month mortality associated with ambulatory cardiology care. However, reasonable deviations from the assumptions underlying each method led to estimated differences in mortality ranging from a 6% absolute reduction in mortality to a 2% increase among patients who received cardiology care. Choosing an analytic strategy depends on both available data and the policy question of interest. We believe that comparative analyses such as this one, with extensive assessment of the assumptions underlying each method, can provide valuable insights into important policy questions reliant on the analysis of observational data.

**Keywords:** causal inference, propensity scores, instrumental variables, acute myocardial infarction, specialty care

## 1. Introduction

The quality and outcomes of care provided by primary care physicians and specialists are increasingly important issues in health policy research. For example, previous studies have documented that acute myocardial infarction (AMI) or heart attack patients treated by cardiologists in the inpatient setting were more likely to receive beneficial therapies

Address correspondence to: Mary Beth Landrum, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115-5899. Tel. 617 432 2460; Fax: 617 432 2563;  
E-mail: landrum@hcp.med.harvard.edu

(Ayanian et al., 1997; Borowsky et al., 1995; Jollis et al., 1996; Nash et al., 1997; Frances et al., 2000). In some of these studies cardiologists' patients were also found to have significantly lower adjusted mortality rates (Jollis et al., 1996; Nash et al., 1997), while others found no adjusted effect of specialty care on mortality (Ayanian et al., 1997; Francis et al., 2000). In contrast, relatively little is known about the quality and outcomes of MI care as a function of the specialty of the physicians providing care following discharge from the hospital. We have recently undertaken a study of outpatient specialty care in AMI survivors using a clinically rich data set linked to outpatient administrative data.

As with many important policy analyses, studies of specialty care generally rely on the analysis of observational data. Characteristics of the patients may affect treatment decisions, such as the decision of patients to seek specialty care or the referral of patients to specialty care by general physicians. Thus patients who receive care from a specialist are often different in terms of both observed and unobserved characteristics from patients who do not receive specialty care, potentially biasing estimates of specialty care effects. Previous studies of inpatient cardiology care in fact demonstrated this selection bias, showing that patients receiving inpatient specialty care for their AMI were younger, were less likely to have chronic illnesses, and were more likely to have prior cardiac disease (Frances, 2000). Moreover, estimates of the relationship between inpatient specialty care and patient outcomes have been found to be sensitive to the analytic strategies chosen to estimate the specialty effects (Francis, et al., 2000). Standard analytic methods adjust for observed differences between treatment groups by stratifying or matching patients on a few observed covariates or with regression analysis in the case of many observed confounders. However, if patients who receive specialty care differ greatly in observed characteristics from those who receive only generalist care, estimates of the specialty effect from regression models rely on model extrapolations and the resulting conclusions can be very sensitive to model mis-specification (Rubin, 1979). Design strategies, such as randomized trials or pre-post studies, can also be employed to avoid selection biases. Practical considerations however suggest that these designs are not likely to be applied to studies of specialty care and thus there is a need to examine more sophisticated analytic strategies to estimate causal effects of specialty care.

Propensity score methods (Rosenbaum and Rubin, 1983a, 1984) have been proposed as a less parametric alternative to regression adjustment and are being increasingly used in health policy studies (Connors et al., 1996; Polanczyk et al., 2001; D'Agostino, 1998 and references therein). Regression models and propensity score methods can reduce bias in causal estimates due to observed differences between treatment groups, but are still subject to biases from *unobserved* differences. Instrumental variable (IV) methods (Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996), used extensively by economists and social scientists, have been recently adopted in health policy studies to estimate causal effects in the presence of unobserved differences between treatment groups (McClellan, McNeil and Newhouse, 1994; Frances et al., 2000). This approach involves identifying variables, denoted instruments, that are related to treatment but not to outcomes other than through their effects on treatment. For example, Frances et al. (2000) recently employed the differential distance between a patient's residence and the nearest hospital in which a majority of patients received cardiology care as an instrument for receipt of inpatient specialty care.

In this paper we illustrate and compare two different analytic approaches to infer the causal effect of ambulatory cardiology care on mortality following an AMI. First, we employed propensity score methods to control for a large number of observed differences between patients who received ambulatory cardiology care and those who received only generalist care. We compared causal estimates from the propensity score approach to an instrumental variable approach in which we employed the density of cardiologists in the patient's county of residence as an instrument for ambulatory cardiology care. We focused on methods that are easy to implement in practice and that require minimal parametric assumptions. Moreover, we considered in depth the assumptions underlying each approach and tested the sensitivity of causal estimates to violations of key assumptions.

There are several advantages to using ambulatory specialty care as a model for comparing these two analytic strategies. First, with the increasing presence of managed-care organizations and limited access to specialty care, understanding the proper mix of generalist and specialty care after patients are hospitalized for serious illnesses such as AMI is increasingly important from a policy perspective. In addition, estimating the effect of specialty care is similar to other important health policy analyses that are subject to observed and unobserved selection biases and unlikely to be studied with a randomized trial. Finally, we had a large representative data set with which to study ambulatory cardiology care containing both extensive clinical detail and a plausible instrument for specialty care. The objective of the paper is both to illustrate the two methods using an important health policy data set and to discuss the analytic issues and choices inherent in policy analyses using observational data. We conclude that comparative analyses using several analytic approaches with extensive assessment of the assumptions underlying each method can provide valuable insights into important policy questions reliant on the analysis of observational data.

In Section 2 we describe the application and data used to estimate the causal effect of outpatient cardiology care. We discuss analytic methods for causal inference framing the problem in the context of the Rubin Causal Model (Rubin, 1974, 1978; Holland, 1986), and describe our analyses in Section 3. In Section 4 we present the results of our analyses including sensitivity analyses of the robustness of the causal estimates to violations of critical assumptions. We conclude with a discussion of differences in the approaches and limitations of the analyses in Section 5.

## **2. Ambulatory Specialty Care Following an AMI**

We obtained our data through the Cooperative Cardiovascular Project (CCP), an initiative of the Health Care Financing Administration (HCFA) to improve care for Medicare patients (Marciniak et al., 1998). This initiative collected data on a national cohort of over 200,000 Medicare patients treated for an AMI in 1994 and 1995. Patients were identified from administrative records based on a discharge diagnosis of AMI (ICD-9 code of 410.xx, excluding patients with 410.x2). Detailed clinical data, including symptoms at presentation, prior medical history, diagnostic and laboratory test results, and the use of

diagnostic and therapeutic procedures during the hospitalization, were then abstracted from patients' medical records.

We studied fee-for-service patients treated in the state of New York. We restricted our sample to fee-for-service beneficiaries because we relied on billing data to identify the specialty of physicians providing care in the outpatient setting that were not available for managed-care patients. In preliminary analyses of seven states, we determined that some of the assumptions required to use density of cardiologists as an instrument for specialty care held best in patients treated in New York (see Section 3.2.1) and thus restricted our sample to these patients. Because we studied outpatient specialty care, we included only patients who survived at least 90 days following discharge and thus had a reasonable opportunity to see a cardiologist. In addition, because specialty care is rarely used for the very elderly or for patients with severe comorbidities, we excluded patients over the age of 85, patients with a documented do not resuscitate order during the index hospitalization and long-term nursing home residents. Finally we excluded patients without clinical confirmation of an AMI in their medical records (Marciniak et al., 1998).

We then linked the abstracted medical record data to Medicare Physician/Supplier (Part B) Files and Hospital Outpatient Files to determine the specialty of physicians providing outpatient care in the 90 day period following discharge. Our primary comparison was the receipt of any cardiology care versus the receipt of care only from generalists. We thus categorized patients meeting our inclusion criteria into two treatment groups: a) generalist care alone: 1916 patients (35%) who had at least one office visit with a primary care physician (family practitioners, general practitioners or internists) without any office visits with a cardiologist and b) cardiology care: 3551 (65%) who had at least one office visit with a cardiologist. We excluded a small number of patients (421) who received only non-cardiology specialty care or no ambulatory care.

Our outcome variable was mortality at 18 months after AMI discharge, obtained from the Health Insurance Master File. We obtained structural characteristics of the hospitals through the HCFA Provider of Service File, the American Hospital Association Survey and a telephone survey. We identified patients who received cardiology care during their index hospitalization using the Medicare Physician/Supplier (Part B) files. Patients were coded as having received inpatient cardiology care if either the admitting physician was a cardiologist or if they received at least one cardiology consult during their hospitalization. Finally, we computed the density of cardiologists in each patient's county of residence using data from the Area Resource File (Bureau of Health Professionals). Specifically we computed density as the number of practicing cardiologists in 1994 divided by the estimated county population in 1994 aged 65 and over.

We report demographic, clinical, and provider characteristics in each of the two treatment groups in the first section of Table 1 labeled "Full Sample". Unadjusted mortality was substantially lower at 180 days (9.3% versus 15.8%) among patients who received ambulatory cardiology care. However, this section of Table 1 demonstrates that the use of specialty care in the ambulatory setting was associated with observed characteristics of the patients. Younger patients with less co-morbid disease, and patients with prior cardiac disease or those with a more severe MI (patients with recurrent chest pain, shock or cardiac arrest) were more likely to receive ambulatory cardiology care.

Table 1. Observed characteristics according to specialty care.

	Full Sample			Matched Sample			Density of Cardiologists		
	Cardiology Care (n = 1916)	Generalist Care (n = 3551)	Stand Diff. <sup>†</sup>	Cardiology Care (n = 1775)	Generalist Care (n = 1775)	Stand Diff. <sup>†</sup>	> 6.7 per 10,000 <sup>†</sup> (n = 2597)	< 6.7 per 10,000 <sup>†</sup> (n = 2870)	Stand Diff. <sup>†</sup>
<i>Demographic Characteristics</i>									
Age	73.4	74.5	-19.8	74.2	74.3	-0.7	73.9	73.7	4.3
Male	57.3	50.6	13.5	52.5	51.7	1.6	54.8	55.1	-0.6
Race									
White	93.7	91.3	9.1	91.2	91.8	-2.4	92.6	93.1	-2.2
Black	3.1	4.6	-7.9	4.6	4.2	1.8	3.9	3.4	3.0
Hispanic	2.5	3.1	-4.0	3.4	3.2	1.4	2.7	2.7	-0.3
Other	0.7	0.9	-2.6	0.9	0.8	0.6	0.8	0.8	0.3
<i>Co-Morbidities</i>									
Stroke	7.6	9.9	-8.1	9.1	9.5	-1.4	8.2	8.6	-1.4
PVD	8.7	9.8	-3.8	9.6	9.6	0.0	9.3	8.9	1.5
COPD	15.6	20.5	-12.9	19.1	19.1	0.0	16.7	17.9	-3.2
Diabetes	30.1	35.0	-10.5	32.9	34.3	-2.9	30.9	32.8	-4.0
Dementia	1.0	2.0	-8.0	1.5	1.5	0.0	1.4	1.3	1.3
Hypertension	63.4	65.4	-4.2	65.1	65.1	0.0	63.4	64.8	-3.1
Independent Mobility	82.3	77.8	11.3	79.3	78.6	1.7	81.5	79.9	4.0
<i>Prior Cardiac Disease</i>									
CHF	14.7	18.5	-10.2	16.6	17.4	-2.0	16.1	16.1	0.0
AMI	28.7	28.0	1.5	27.7	27.7	0.0	27.2	29.7	-5.5
Angina	58.5	54.9	7.3	55.3	55.3	0.1	54.9	59.8	-9.9
Bypass surgery	13.0	10.2	8.9	10.3	10.6	-1.2	11.7	12.4	-2.0
Angioplasty	6.0	3.9	9.6	4.5	4.1	2.1	4.8	5.8	-4.5

Table 1 continued

<i>Cardiac Events during Hospitalization</i>										
Recurrent chest pain	29.0	26.4	6.0	27.0	27.1	-0.1	27.7	28.5	-1.7	
Shock	2.8	2.3	3.0	2.6	2.4	1.4	2.8	2.3	3.0	
Cardiac arrest	5.2	3.8	6.7	3.8	4.1	-1.1	5.1	4.4	3.4	
CHF	37.1	41.0	-7.9	41.6	40.5	2.3	38.8	38.1	1.4	
<i>Hospital Characteristics</i>										
Rural	4.6	10.2	-21.4	8.1	8.1	-0.2	1.5	12.3	-41.5	
Teaching	70.8	56.8	29.5	61.3	59.5	3.7	70.6	60.7	20.8	
Invasive cardiac services	61.4	49.0	25.3	52.2	51.1	2.2	65.0	48.3	33.9	
<i>Treatments during Hospitalization</i>										
Cardiology care	74.6	53.9	44.1	58.3	57.8	1.0	70.1	64.2	12.6	
Angiography	33.8	20.6	30.1	22.4	22.1	0.6	35.4	22.4	29.4	
Angioplasty	8.0	4.3	15.2	4.6	4.7	-0.2	8.3	4.9	14.3	
Bypass surgery	7.7	5.0	11.2	6.2	5.4	3.2	8.5	4.9	14.5	

Observed covariates in patients stratified by observed treatment in the full sample, in patients stratified by observed treatment in sample of patients matched according to estimated propensity score, and in patients stratified by the density of cardiologists in their county of residence.

\*Standardized difference was calculated as  $100 * (\bar{x}_1 - \bar{x}_0) / \sqrt{((s_1^2 - s_0^2)/2)}$  where  $\bar{x}_1$  and  $\bar{x}_0$  are the sample means in the cardiology care and generalists care treatment groups respectively and  $s_1^2$  and  $s_0^2$  are the sample variances.

‡Median density of cardiologists in the patient's county of residence.

PVD = peripheral vascular disease, COPD = chronic obstructive pulmonary disease, CHF = congestive heart failure, AMI = acute myocardial infarction.

Moreover, patients treated by cardiologists in the ambulatory setting were more likely to have been admitted to urban teaching hospitals with invasive cardiac capabilities, to have received invasive cardiac therapies, and to have been treated by a cardiologist during their hospitalization. Adjustment of these observed differences with a logistic regression model reduced the unadjusted absolute differences in 18-month mortality between treatment groups from 6.5% to 2.9%. However, when treatment groups differ greatly in observed characteristics, estimates of the specialty effect from regression models rely on model extrapolations and the resulting conclusions can be very sensitive to model mis-specification (Rubin, 1979). In addition, given the differences in observed characteristics, it is also likely that patients treated by cardiologists in the outpatient setting differed in *unobserved* characteristics from those who received care only from generalists. For example, while we observed a large number of important clinical events both prior to and during the hospitalization, the CCP did not collect data on clinical events following hospitalization that may have affected decisions to seek (or be referred to) specialty care. We next discuss our analytic strategies for obtaining estimates of the causal effect of ambulatory specialty care on subsequent patient mortality in the presence of these observed and potentially unobserved differences between treatment groups.

### 3. Analytic Methods for Causal Inference

The Rubin Causal Model (Rubin, 1974, 1978; Holland, 1986) frames the estimation of causal effects as a comparison of potential outcomes under differing treatments. We restrict our attention to a binary treatment such as ambulatory specialty care. Let  $D_i$  represent a binary variable equal to 1 if patient  $i$  received the treatment under study and  $Y_i(D_i)$  be the potential outcome of patient  $i$  given treatment.  $Y_i(1)$  represents the outcome that would have been observed had the patient received the treatment, while  $Y_i(0)$  is the outcome that would have been observed had he or she received the control. The causal effect of D on Y for patient  $i$  is defined to be  $Y_i(1) - Y_i(0)$ . We only observe one of these outcomes for each patient and thus cannot directly observe the causal effects. However, under certain assumptions regarding treatment assignment, we can estimate the average causal effect,  $E(Y(1) - Y(0))$ . For example, in a randomized experiment with perfect compliance, observed treatment is independent of potential outcomes so that the difference in mean responses between the two treatment groups is an unbiased estimate of the average causal effect.

In the absence of randomization, observed treatments are generally not independent of potential outcomes and moreover, the treatment assignment mechanism is neither in control of the experimenter nor completely observed. In the absence of randomization, propensity score and instrumental variable methods can lead to unbiased estimates of causal effects under differing sets of assumptions regarding the assignment of treatment. In the following sections we discuss our application of these two methods to estimate the average causal effect of ambulatory specialty care. Because the mechanisms by which patients were assigned to receive specialty care were not completely observed, the assumptions underlying each method are inherently untestable. We consider the plausibility of the assumptions underlying each method using observed data and clinical

knowledge and also describe methods to test the sensitivity of causal estimates to violations of these critical assumptions.

### *3.1. Propensity Score Methods*

The propensity score approach, which involves comparing patients with similar propensity to receive cardiology care, attempts to balance observed characteristics in the treatment groups as would occur in a randomized experiment. The propensity score for patient  $i$ ,  $e(x_i) = \Pr(D_i = 1 | X_i = x_i)$ , is defined to be the probability of receiving treatment given a vector of observed characteristics,  $X_i = x_i$ . Rosenbaum and Rubin (1983) have shown that stratifying or matching on the propensity score produces treatment groups that are balanced in terms of the observed variables contained in  $X$ . Thus, under the assumption that treatment assignment is ignorable given  $X$ , that is treatment assignment,  $D$ , and potential outcomes,  $(Y(1), Y(0))$ , are conditionally independent given  $X$ , unbiased estimates of the average causal effect can be obtained by comparing patients with similar values of the propensity score.

#### ***3.1.1. Application of Propensity Score Methods in the Ambulatory Care Study***

Application of propensity score methods involves three steps: (1) estimation of the propensity score followed by (2) matching or stratification of patients according to their estimated propensity score and (3) comparison of outcomes in matched or stratified patients.

To estimate the propensity score, we fit a logistic regression model for the receipt of any cardiology care in the 90 days following discharge as a function of patient demographic and clinical characteristics, treatments received during hospitalization, and characteristics of the inpatient providers and hospitals. The final model contained 41 clinical and provider characteristics, including all of the characteristics reported in Table 1. We did not include the density of cardiologists in the county in the propensity score model. While we expect this variable to predict receipt of cardiology care, we do not expect patient outcomes to vary according to the density of cardiologist in their area (in fact this is one of the assumptions required by the instrumental variable analysis, see Section 3.2.1) and thus it was not important to balance treatment groups in terms of this covariate.

The predicted probabilities of receiving cardiology care in the two treatment groups are shown in Figure 1 and provide several useful diagnostics. First, we can use the average estimated probability of receiving specialty care in the two treatment groups to summarize the observed differences in the 41 covariates contained in the propensity score model. There was an absolute difference of 11% in the average predicted probability of receiving cardiology care (69% in the cardiology care group compared to 58% among patients receiving only generalist care), suggesting moderate differences in the observed characteristics between the two treatment groups. Also, with the exception of a few generalist patients with very low predicted propensity to receive specialty care, there was substantial overlap in estimated propensity scores. Overlap between the two treatment groups is necessary for creating comparison groups with similar observed characteristics by matching or stratifying on the estimated propensity score. A lack of overlap would



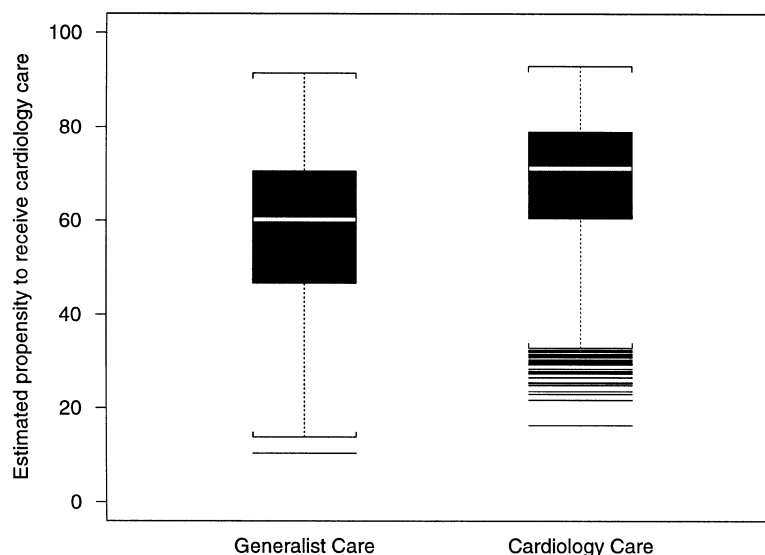


Figure 1. Estimated propensity scores. Boxplots displaying the distribution of the predicted probability of receiving cardiology care in patients stratified by observed treatment.

imply that there were combinations of covariate values found in only one of the two treatment groups so that patients who received cardiology care cannot be meaningfully compared to those receiving generalist care only. Finally, 34 patients (displayed as outliers in Fig. 1) received cardiology care despite a very low predicted propensity score, suggesting we may not have observed all predictors of specialty care. The existence of unobserved variables related to treatment choice places the assumption of an ignorable treatment assignment mechanism in doubt. We consider sensitivity to possible unobserved confounders in Section 3.1.3.

To perform a stratified analysis, we divided patients into quintiles according to their estimated propensity score. We then fit a series of ANOVA models to examine the balance between treatment groups within quintiles in each of the covariates. The dependent variable in each model was the covariate of interest and the models included main effects for treatment group and quintile of estimated propensity score as well as interactions between treatment and propensity score quintile. Tests of the treatment main effects and interactions between treatment and propensity score quintiles provide evidence of imbalance in the covariates between treatment groups after stratification on the estimated propensity score. Four examples of covariates in patients stratified by treatment status and quintile of propensity score are displayed in Figure 2. There were no significant differences between treatment groups in these four covariates or in any of the covariates contained in the propensity score model after stratifying patients into quintiles of estimated propensity score. In addition, there were only three significant interactions between treatment and quintile of propensity score in the ANOVA models, pointing to residual imbalance in the number of patients treated in rural hospitals, in the number of hypertensive patients, and in

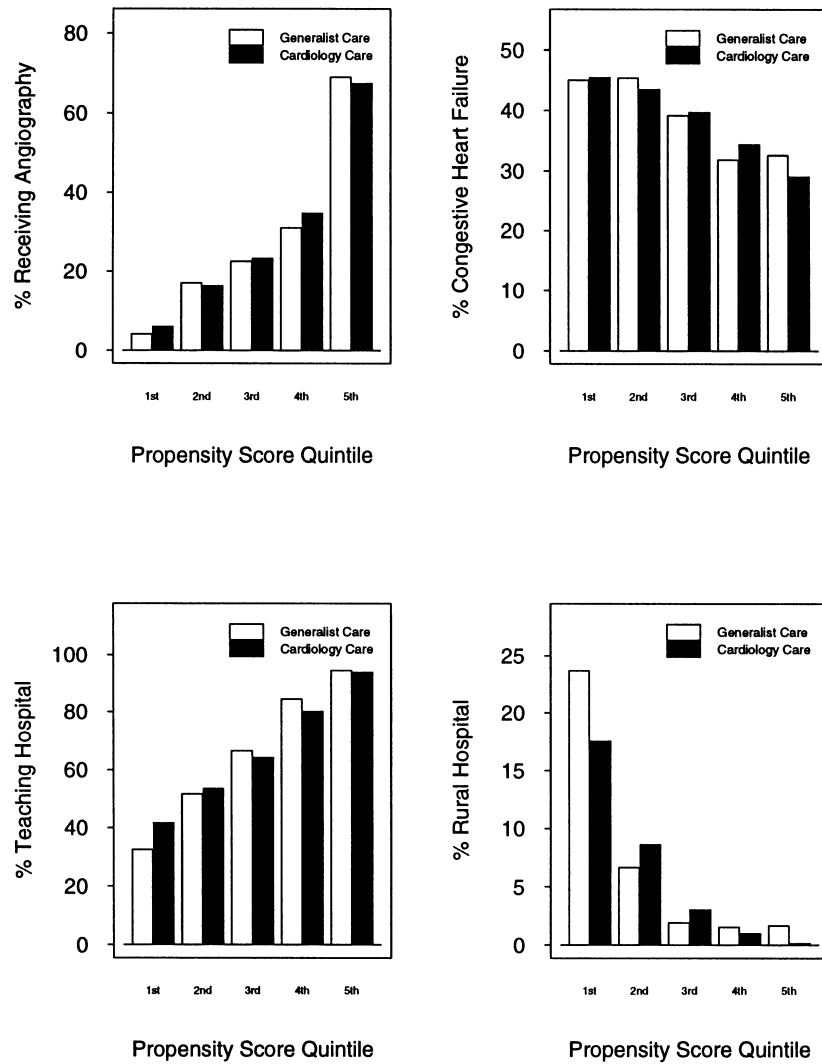


Figure 2. Balance of selected covariates. Covariate means in patients stratified by estimated propensity score and observed treatment.

the number of patients experiencing a cardiac arrest during the hospitalization in some of the strata. Thus we concluded that stratifying patients according to their estimated propensity to receive cardiology care removed most of the bias in observed characteristics between the two treatment groups.

Under the assumption that the observed characteristics contained in the propensity score model were the only variables related to both receipt of specialty care and patient

outcomes, the difference in mortality rates between the two specialty groups within each stratum is an unbiased estimate of the causal effect of ambulatory specialty care in that stratum. Estimates of the overall causal effect of specialty care on mortality were then obtained by combining the stratum-specific differences using standard methods for direct adjustment of a covariate (Rosenbaum and Rubin, 1984). We tested the sensitivity of the analyses to the choice of the number of strata, repeating the above analyses after stratifying patients into 3 and 10 equal-sized groups, respectively, according to their estimated propensity score.

Figure 2 demonstrates that while stratifying patients according to their estimated propensity to receive cardiology care removed much of the observed differences between treatment groups, there was still residual imbalance in some of the variables, particularly in the lowest and highest propensity strata. To further reduce the differences in observed characteristics, we matched patients according to their estimated propensity to receive cardiology care. To obtain a matched sample, we attempted to match each generalist patient to the cardiologist patient with the closest estimated propensity to receive cardiology care within a caliper of propensity scores chosen so that we may expect an approximate 90% reduction of the differences between treatment groups (Rosenbaum and Rubin, 1985a). Specifically, we randomly selected a generalist patient and attempted to match him or her to the cardiology patient with the closest estimated propensity score (on the logit scale) within 0.6 times the pooled standard deviation of estimated logits. If no cardiology patient could be found within the propensity score caliper, we removed the generalist patient from the sample. Using this method, we successfully matched 1775 of the 1916 generalist patients (93%), leaving 1776 of the 3551 cardiology patients unmatched. Unmatched cardiology patients were those with the highest estimated propensity to receive cardiology care. In particular, they were younger, had less comorbid disease, were more likely to have prior cardiac disease, and were more likely to be admitted to urban, teaching hospitals with invasive cardiac capabilities. Note that this approach attempts to achieve a compromise between bias introduced by observed differences between treatment groups and bias introduced by analyzing a non-representative sample of treated patients (Rosenbaum and Rubin, 1985b).

We report differences in observed characteristics between specialty groups in the matched sample in the middle Section of Table 1, labeled “Matched Sample”. These results demonstrate the power of matching on the estimated propensity score to balance treatment groups in terms of observed characteristics. In the matched sample, the standardized differences in covariate means were all smaller than 4% and substantially reduced compared to the original differences between treatment groups. We estimated the average causal effect of specialty care by comparing mortality rates in the two specialty groups in the matched sample. Under the assumption that all variables related to both treatment and outcome were included in the propensity score model, this analysis will provide an unbiased estimate of the expected benefit of ambulatory specialty care.

### ***3.1.2. Assessment of Propensity Score Assumptions***

Propensity score methods lead to unbiased estimates of causal effects under the assumption that receipt of specialty care was independent of patient outcomes, conditional on a set

of observed characteristics. This implies that conditional on all of the clinical, hospital, and treatment variables contained in the propensity score model, cardiology care in the outpatient setting was randomly assigned. This assumption would be violated if unobserved characteristics of the patients independently impacted the likelihood of receiving cardiology care. Potential unobserved confounders in this study include clinical course after discharge, measures of disease severity observed by the providers but not abstracted from the medical records, and unobserved characteristics of the patients such as social support that led them to seek specialty care.

### ***3.1.3. Sensitivity of Propensity Score Estimates to Violations of Assumptions***

We examined the robustness of the causal estimates to unobserved differences between treatment groups using methods proposed by Rosenbaum and Rubin (1983b). Unobserved confounders can bias propensity-score causal estimates in either direction. For example, if patients with continued cardiac symptoms after discharge were more likely to receive ambulatory cardiology care and had increased mortality compared to patients without continued complications, propensity-score estimates of the average causal effect will be smaller than the true effect of cardiology care. Alternatively, if patients with unobserved co-morbidities were both less likely to receive specialty care and had higher mortality, propensity score estimates will be too large. We examined the sensitivity of our conclusions to both of these potential effects by hypothesizing the existence of an unobserved binary variable related to both the receipt of cardiology care and mortality.

We considered hypothetical confounding effects of similar magnitude to those in the observed data. For example, in fitting the propensity score model we observed estimated odds-ratios of receiving cardiology care ranging from 0.6 (patient discharged to a nursing home) to 2.5 (patient received inpatient cardiology care). Similarly, estimated effects in a model for 18-month mortality ranged from 0.4 (patient received bypass surgery) to 2.7 (metastatic cancer). Thus we considered situations in which the hypothetical variable decreased the odds of receiving cardiology care from 33% to 67% while increasing the odds of mortality from 50% to 300%. Similarly we considered a hypothetical confounder that increased both the odds of cardiology care and mortality from 50% to 300%. Updated estimates of the average causal effect after adjustment for the unobserved confounder were then obtained using the results from the stratified analysis for each set of assumptions regarding the prevalence of the unobserved confounder and its relationship to mortality and specialty care. Updated effects under a range of such assumptions provide a range of estimated causal effects that plausibly could have been found had we been able to adjust for the hypothetical confounder.

### ***3.2. Instrumental Variable Methods***

The propensity score approach is a powerful method for balancing treatment groups in an observational study according to *observed* characteristics, but only controls for *unobserved* characteristics to the extent that they are correlated with the observed variables. Alternatively, instrumental variable methods seek to estimate causal effects in the presence of

unobserved differences between treatment groups. This approach involves identifying variables, denoted as instruments, that are related to treatment but not to outcomes except through their effects on treatment. Application of IV methods typically proceeds in two stages. First, the instruments are used to predict treatment independent of unobserved selection effects. Patient outcomes are then compared in terms of predicted treatment rather than actual treatment.

Let  $Z_i$  denote a vector of instrumental variables. In our case  $Z_i$  will be a vector of variables describing the density of cardiologists in the patient's county of residence. For example, if we stratify patients according to the median density of cardiologists in the county,  $Z_i$  would be a binary variable equal to 1 if the patient resided in a county with more than 6.7 cardiologists per 100,000 estimated elderly population. The potential outcome framework can be extended to allow for potential treatments as a function of  $Z$ . Let  $D_i(Z_i)$  denote the treatment status of patient  $i$  conditional on  $Z_i$ . With a binary instrument,  $D_i(1)$  represents the treatment that would have been observed if the patient lived in an area with a high density of cardiologists, while  $D_i(0)$  is the treatment he or she would have received had he or she lived in a low-density area.

With unobserved differences between treatment groups, the average causal effect of treatment cannot be estimated without bias even under a strict set of assumptions (described in detail in Section 3.2.1). However, Imbens and Angrist (1994) showed that under the set of assumptions described below, a Local Average Treatment Effect (LATE) can be estimated for the subset of patients for whom the instrument determines treatment, i.e. those for whom  $D_i(1) > D_i(0)$ . These patients, called "compliers" (Angrist, Imbens and Rubin, 1996) or "marginal patients" (Harris and Remler, 1998; McClellan, McNeil and Newhouse, 1994), are those who would receive cardiology care if they lived in an area with a high density of cardiologists but not if they lived in a low-density area. Note that because we observe only one of the potential treatments, we cannot identify compliers using observed data.

In the case of a binary instrument, the local average treatment effect is estimated as (Angrist, Imbens, and Rubin, 1996):

$$E(Y_i(1) - Y_i(0) | D_i(1) > D_i(0)) = \frac{\bar{Y}_{Z=1} - \bar{Y}_{Z=0}}{\bar{D}_{Z=1} - \bar{D}_{Z=0}} \quad (1)$$

In equation 1,  $\bar{Y}_{Z=1}$  and  $\bar{Y}_{Z=0}$  are the mean outcomes among patients with  $Z_i = 1$  and  $Z_i = 0$  respectively, and  $\bar{D}_{Z=1}$  and  $\bar{D}_{Z=0}$  are the fraction of patients who received treatment among patients with  $Z_i = 1$  and  $Z_i = 0$ , respectively. Intuitively, under the assumptions that patients differing according to the instrumental variable do not differ in terms of health characteristics (both observed and unobserved) and that  $D$  is the only treatment determined by the instrument that impacts outcomes, we can directly compare outcomes in patients in the two instrument groups. Scaling the outcome differences by differences in treatment status induced by the instrumental variable then produces an unbiased estimate of the causal effect of  $D$  on the subset of patients for whom the instrument determined treatment.

### 3.2.1. Assessment of IV Assumptions

Obtaining unbiased estimates of causal effects using instrumental variables relies on several key assumptions as outlined by Angrist, Imbens and Rubin (1996).

- *Stable Unit Treatment Value Assumption.* First, potential treatments and outcomes for each patient are assumed to be unrelated to the treatment status of all other patients. This assumption implies that receipt of specialty care and mortality were not affected by the specialty of physicians providing ambulatory care to other patients. While this seems plausible in this example, to the extent that access to providers varies across geographic areas, this assumption may be violated. For example, patients living in areas with a high density of cardiologists may be more likely to have increased access to all kinds of specialists, while treatment by a generalist may be more likely in areas where other patients are also receiving care from generalists.
- *Non-zero causal effect of instrument on treatment.* Second, instrumental variables must predict treatment status. We computed the observed fraction of patients who received cardiology care among patients stratified into approximate quintiles according to the density of cardiologists in their county of residence. The likelihood of receiving cardiology care was significantly and positively associated with quintile of cardiology density (p-value < 0.001), increasing from 57% to 72% as the density of cardiologists in the county increased from less than 4 to more than 12 per 100,000, providing empirical support for this assumption. We found, however, that density of cardiologists was not associated with receipt of ambulatory specialty care in several other states (Texas, California and Massachusetts), suggesting that an instrumental variable analysis using density of cardiologists would not be appropriate with data from these states.
- *Ignorable assignment of the instrument.* This assumption implies that patients differing according to the density of cardiologists in their area were similar in terms of observed and unobserved characteristics as would have occurred if density of specialists in the county was randomly assigned. While this assumption cannot be verified directly, examination of observed characteristics as a function of the instrument can provide some evidence of its validity. In the last section of Table 1, labeled “Density of Cardiologists”, we report patient characteristics in two groups of patients differing in the density of cardiologists in their county of residence. Patients stratified according to the density of cardiologists in their area were very similar in demographic and clinical characteristics (standardized differences were all less than 10% and most were less than 5%), providing evidence in favor of ignorable assignment of the instrument. We also examined the balance of observed characteristics between patients grouped according to quintiles of the instrument and found results similar to those reported in Table 1. However, there was evidence of differences in observed characteristics in patients stratified according to the density of cardiologists in their county of residence in several other states. For example, patients living in areas with a high density of cardiologists in Florida tended to be older than those living in areas with fewer specialists.
- *Exclusion restriction.* Fourth, the instrument is assumed to have no effect on outcomes other than through its effect on treatment. While this assumption also cannot be verified directly, results reported in the last section of Table 1 place this assumption in doubt. In

particular, the density of cardiologists in an area was correlated with hospital characteristics (such as teaching status and the availability of invasive cardiac services) and treatments that the patient received during hospitalization (such as coronary angiography and inpatient cardiology care). Thus the instrument may have an effect on patient mortality through these other treatments. Moreover, as seen in Table 1, areas with a high density of cardiologists tended to be more urban and processes of care may differ between urban and rural areas. To the degree that these violations are observable (as is the case for the variables contained in Table 1), they can be controlled for in the analysis. The potential violation of this critical assumption in this example is addressed both with stratified analyses described in Section 3.2.2 and with sensitivity analyses described in Section 3.2.3.

- *Monotonicity.* Finally, the instrument is assumed to affect treatment monotonically. This assumption implies that if a patient in a low-density area received cardiology care he or she would also have received cardiology care if he or she lived in a high-density area. Again, this assumption cannot be directly verified but seems reasonable in this example.

### 3.2.2. Application of IV Methods in the Ambulatory Care Study

We employed IV methods to estimate the impact of ambulatory specialty care on patient mortality by first dividing patients into approximate quintiles according to the density of cardiologists in their county of residence. In the case of a multivalued instrument, the local average treatment effect can be obtained by averaging pairwise effects such as those estimated with equation 1 (Imbens and Angrist, 1994) into an overall estimate of the treatment effect among patients for whom the supply of cardiologists determined treatment. Because hospital characteristics and inpatient treatments were correlated with the supply of cardiologists, we also estimated local average treatment effects at fixed values of some of these characteristics. We estimated the causal effect of specialty care within four groups of patients categorized according to whether they were admitted to a teaching hospital and whether they received coronary angiography during the hospitalization. The cell sizes ranged from 241 (patients admitted to a non-teaching hospital who received angiography) to 2246 (patients admitted to a teaching hospital who did not receive angiography). Overall estimates of the local average treatment effect were then obtained as a weighted average of the estimated effects within each cell. Ideally we would have stratified on additional observed patient, hospital and inpatient treatment characteristics, but were limited by the available sample size. An alternative approach would have been to use parametric models for the relationships between the covariates, treatment, and outcomes. (For examples of instrumental variable analyses with such modeling assumptions, see Hirano et al., 2000 and Glickman and Normand, 2000). We instead focused on non-parametric approaches that did not require modeling assumptions beyond the required assumptions regarding treatment assignment. Finally, because density of cardiologists was strongly related to urbanicity, we estimated causal effects separately in 5106 patients admitted to an urban hospital.

### 3.2.3. Sensitivity of IV Estimates to Violations of Assumptions

One of the key assumptions of IV analyses, the exclusion restriction, asserts that density of specialists affected patients' outcomes only through its effect on the receipt of specialty care. The observed relationships between cardiology supply and hospital and inpatient treatment characteristics reported in Table 1 demonstrate potential violations of this assumption. We attempted to control for some of these observed differences using the stratified analyses described above. We also tested the sensitivity of the IV estimates to unobserved violations of the exclusion restriction using methods proposed by Angrist, Imbens and Rubin (1996).

Let  $Y_i(Z_i, D_i(Z_i))$  represent the potential outcome for subject  $i$  given observed set of instrumental variables  $Z_i$  and potential treatment status,  $D_i(Z_i)$ . The exclusion restriction asserts that  $Y_i(Z_i = z, D_i(Z_i = z)) = Y_i(Z_i = w, D_i(Z_i = w))$ . In our study, this assumption implies that conditional on the specialty of the physicians providing ambulatory care, if a patient would have died within 18-months of discharge had he or she lived in an area with a low density of cardiologists, then he or she also would have died had he or she lived in an area with a high density of cardiologists. Angrist, Imbens and Rubin (1996) showed that if a binary instrument exerts an additive effect on outcomes conditional on the treatment of interest due to the instrument's association with other correlated "treatments", the bias in IV estimates is equal to:

$$\frac{E(Y_i(1, d) - Y_i(0, d))}{E(D_i(1) - D_i(0))} \quad (2)$$

where  $Y_i(1, d)$  is the outcome of patient  $i$  if  $Z_i = 1$  and he or she received treatment  $d$ ,  $d = 0, 1$  and similarly,  $Y_i(0, d)$  is the outcome under  $Z_i = 0$ . We tested the sensitivity of the IV estimates by considering a hypothetical binary treatment associated with both the density of cardiologists in the area and a reduction in patient mortality. The numerator in equation 2 represents the additional effect of living in an area with a high density of cardiologists conditional on the specialty of the physicians providing ambulatory care due to the increased likelihood of receiving the hypothetical treatment. The denominator in equation 2, the probability of being a complier (a patient who would receive specialty care if he or she lived in an area with a high density of cardiologists but not if he or she lived in an area with few cardiologists), can be estimated using the observed data. We computed the expected bias using equation 2 under a variety of assumptions regarding the association between density of cardiologists and the hypothetical treatment and the effect of the hypothetical treatment on patient mortality.

## 4. Results

### 4.1. Propensity Score Methods

Propensity score estimates of the average causal effect of specialty care are reported in Table 2. In the first 5 rows we report 18-month mortality in the two treatment groups and estimated causal effects of specialty care within groups of patients stratified according to



Table 2. Propensity score estimates of average causal effect.

	Cardiology Care		Generalist Care		Average Causal Effect	
	<i>N</i>	18 Month Mortality (%)	<i>N</i>	18 Month Mortality (%)	Difference in Mortality (%)	Standard Error
<i>Stratified Analysis</i>						
Quintile 1 (lowest propensity)	455	12.3	638	23.8	− 11.5	(2.3)
Quintile 2	629	13.5	465	15.9	− 2.4	(2.2)
Quintile 3	725	11.0	368	12.2	− 1.2	(2.1)
Quintile 4	830	7.8	264	9.1	− 1.3	(2.0)
Quintile 5 (highest propensity)	912	4.8	181	3.9	1.0	(1.6)
Overall effect (equally weighted) <sup>§</sup>		9.9		13.0	− 3.1	(0.9) <sup>†</sup>
Overall effect (weighted to cardiology patients)		9.3		11.5	− 2.2	(0.9) <sup>†</sup>
<i>Matched Analysis</i>						
Matched Sample	1775	11.7	1775	14.8	− 3.0	(1.1) <sup>‡</sup>

Estimated difference in 18-month mortality between patients who received ambulatory cardiology care compared to those who received only generalist care using propensity score methods

<sup>†</sup>Standard errors calculated using methods for directly standardized rates.

<sup>‡</sup>Standard errors calculated using methods for matched samples.

<sup>§</sup>Stratifying patients into 3 and 10 equal-sized groups resulted in estimated differences between treatment groups equal to − 3.7% and − 3.1% respectively.

quintiles of estimated propensity scores. Among patients least likely to receive treatment (quintile 1), cardiology care was estimated to reduce absolute 18 month mortality by 11.5% (12.3% versus 23.8%), a significantly larger difference than the estimated causal effect in the other 4 quintiles. In fact, among patients most likely to receive cardiology care (quintile 5), mortality was higher in the cardiology care group, although not significantly so. Estimates of the overall impact of specialty care on mortality were obtained by averaging the stratum-specific differences (Rosenbaum and Rubin, 1984) and are reported in rows 6 and 7. The simple average of the stratum-specific differences estimated mortality to be 3.1% lower among patients who received cardiology care. As each stratum contained an equal number of patients, this estimates the reduction in mortality that we could expect if all patients represented by the AMI cohort received cardiology care. Weighting the stratum-specific differences by the number of cardiology patients in each quintile estimated a 2.2% absolute reduction in 18-month mortality among the patients who typically received cardiology care (sometimes called the average causal effect among the treated). Again, this reduction in the estimated causal effect suggests that patients most likely to benefit from specialty care did not receive it in practice. We also tested the sensitivity of the stratified analysis to the number of strata and found similar results using either 3 or 10 equal-sized groups.

Comparing 18-month mortality in the matched sample (last row of Table 2), we estimated that patients who received ambulatory cardiology care had a 3.0% absolute reduction in mortality at 18 months. This effect is similar in magnitude to the estimate from the stratified analysis, but was estimated less precisely due to the decreased number of patients in the matched sample.

#### 4.1.1. Sensitivity Analyses

Propensity score analyses depend on the assumption of an ignorable treatment assignment mechanism given the observed characteristics in the propensity score model. Our study of ambulatory specialty care had a large number of detailed clinical variables and hospital characteristics to include in the model, making this assumption more plausible. However, we cannot exclude the possibility that unobserved characteristics of the patients independently affected the likelihood of receiving cardiology care.

We examined the sensitivity of our conclusions to potential unobserved confounding effects by hypothesizing the existence of an unobserved binary variable related to both the receipt of cardiology care and mortality. Table 3 reports updated estimates of the average causal effect after adjustment for the hypothetical confounder, assuming that 50% of the patients exhibited the unobserved variable and under varying assumptions regarding its relationship to cardiology care and mortality. For example, if we controlled for a variable that increased both the odds of receiving cardiology care and the odds of mortality by a factor of three, in addition to the 41 clinical variables contained in the propensity score model, we would have observed a 5.9% absolute reduction in mortality associated with cardiology care (compared to the 3.1% reduction found with the propensity score analysis that excluded this hypothetical variable). Across a range of unobserved confounding effects, causal effects ranging from  $-0.5\%$  (which is less than the standard error of the estimated effect) to  $-6\%$  were consistent with our data. Thus our conclusions were mildly

Table 3. Sensitivity of propensity score causal estimates to violations of assumptions.

Effect of unobserved variable on odds of specialty care (Odds-Ratio)	Effect of unobserved variable on odds of 18-month mortality (Odds-Ratio)	Average causal effect controlling for unobserved variable (% diff.)
3.0	3.0	-5.9
2.5	2.5	-5.1
2.0	2.0	-4.2
1.5	1.5	-3.5
0.67	1.5	-2.7
0.50	2.0	-2.0
0.40	2.5	-1.2
0.33	3.0	-0.5

Estimated difference in 18-month mortality between patients who received cardiology care compared to those who received only generalist care controlling for all variables contained in the propensity score model in addition to a hypothetical unobserved variable. Based on a propensity score analysis estimating the overall effect of specialty care as an equally weighted average of effects in patients stratified according to quintile of estimated propensity score (estimated difference =  $-3.1\%$  in the absence of unobserved confounders)

sensitive to unobserved confounding effects that are extreme, but within the range of observed effects.

#### 4.2. Instrumental Variable Methods

We report results of the instrumental variable analyses in Table 4. First we estimated the local average treatment effect without controlling for any of the observed differences in hospital and inpatient treatment characteristics correlated with the supply of cardiologists (first row). The unadjusted IV analysis estimated a 9.5% absolute reduction in mortality rates associated with the receipt of cardiology care. However, controlling for treatment in a teaching hospital and the receipt of coronary angiography during the index hospitalization reduced this estimate to 1.0%. Because cardiology supply was also higher in urban areas, we repeated these analyses using only the 5106 patients admitted to an urban hospital. Restricting the sample to urban patients resulted in slightly larger, but similar estimates.

Estimated standard errors for the local average treatment effects depend on the ability of the instrument to induce variability in observed treatment. For example, the likelihood of receiving cardiology care increased from 57% to 72% as the density of cardiologists in the county increased from less than or equal to 4 to more than 12 per 100,000. Thus, specialty treatment was determined by cardiology supply in only approximately 15% of the population. This resulted in standard errors for the instrumental variable causal estimates that were much larger than those from the propensity score analysis and thus none of the estimated effects were statistically significant.

##### 4.2.1. Sensitivity Analyses

The instrumental variable estimates reported in Table 4 controlled for the increased likelihood of patients living in areas with a high supply of cardiologists to also be treated in

Table 4. Instrumental variable estimates of the local average treatment effect.

	All Patients (n = 5467)		Patients Treated in Urban Hospitals (n = 5106)	
	LATE (% Diff.)	Standard Error	LATE (% Diff.)	Standard Error
No Covariates	− 9.5	(7.9)	− 11.3	(10.0)
Controlling for hospital and treatment characteristics	− 1.0	(8.4)	− 4.4	(9.0)

Local average treatment effect (LATE), the estimated difference in 18-month mortality between patients who received ambulatory cardiology care compared to those who received only generalist care among the subset of patients for whom cardiology supply determined treatment. We estimated causal effects using standard software for two-stage least squares models (Johnston and DiNardo, 1997) with indicator variables for quintiles of density of cardiologists as instruments for the receipt of cardiology care in the first stage of the model. Models controlling for hospital and treatment characteristics included interactions between the instrument indicator variables, teaching status, and receipt of angiography (Angrist and Imbens, 1995)

teaching hospitals and to undergo coronary angiography during their hospitalization. However, density of cardiologists was associated with other observed inpatient treatments and characteristics of the hospitals and may also have been correlated with unmeasured treatments. In this section we report sensitivity analyses examining the robustness of our conclusions to such potential violations of the exclusion restriction. We tested the sensitivity of our IV estimates using equation 2, by considering a hypothetical binary treatment associated with both the density of cardiologists in the area and a reduction in patient mortality. For example, patients residing in areas with a high density of cardiologists may have been more likely to be treated at high-volume hospitals. Previous research has suggested hospitals providing care to larger number of AMI patients have lower mortality rates (Thiemann et al., 1999). If patients residing in areas with a high density of cardiologists were 5% more likely to be treated in a high-volume hospital and treatment in a high volume hospital was associated with a 2% absolute reduction in mortality, we would expect a 0.1% difference in the mortality rates between two groups of patients stratified according to the density of cardiologists in their area. Comparing use of specialty care in the first and fifth quintiles of density of cardiologists, we estimated that the supply of cardiologists determines ambulatory specialty care for 15% of patients. Then using Equation 2, we estimated the bias in the IV causal effect to be equal to  $.001/.15 = 0.6\%$ . We report the potential bias in IV estimates as a function of the increased likelihood of receiving a hypothetical treatment as a result of living in an area with a high density of cardiologists and the mortality benefit associated with the hypothetical treatment in Table 5.

We considered potential bias in the specialty effects estimated using both the full sample and the sample of patients admitted to an urban hospital. Basing inference on the full sample and controlling for hospital and inpatient treatment characteristics, we estimated that controlling for an additional treatment associated with cardiology supply and patient outcomes would result in estimated causal effects ranging from a 2.2% absolute increase in mortality associated with cardiology care to a 0.7% absolute

Table 5. Sensitivity of instrumental variable causal estimates to violations of assumptions.

Increased use of treatment in high density areas (%)	Absolute mortality benefit of treatment (%)	Full Sample LATE = -1%		Urban Patients LATE = -4%	
		Bias (%)	Corrected LATE (%)	Bias (%)	Corrected LATE (%)
10.0	5.0	3.2	2.2	3.9	-0.1
10.0	2.0	1.3	0.3	1.6	-2.4
10.0	1.0	0.6	-0.4	0.8	-3.2
5.0	5.0	1.6	0.6	2.0	-2.0
5.0	2.0	0.6	-0.4	0.8	-3.2
5.0	1.0	0.3	-0.7	0.4	-3.6

Local average treatment effect (LATE), estimated difference in 18-month mortality between patients who received cardiology care compared to those who received only generalist care among the subset of patients for whom cardiology supply determined treatment, correcting for bias due to a hypothetical treatment correlated with cardiology supply

reduction. Restricting the sample to patients admitted to urban hospitals, corrected estimated causal effects ranged from a 0.1% to a 3.6% absolute reduction in 18-month mortality associated with specialty care.

#### *4.3. Comparison of Approaches*

We illustrated and compared two analytic approaches to estimate the causal effect of ambulatory cardiology care on 18-month mortality. Propensity score analyses indicated a small but significant benefit associated with the receipt of cardiology care following discharge from the hospital. This benefit was concentrated among patients with the lowest propensity to receive specialty care. Point estimates from the instrument variable analyses were also consistent with a small benefit to receiving ambulatory specialty care. However, because the IV causal effects were not precisely estimated, the differences between specialty groups were not found to be statistically significant in the IV-based inferences. Before making policy recommendations regarding the use of ambulatory specialty care in elderly MI survivors, several issues should be considered.

First, both propensity score and instrumental variable approaches rely on critical assumptions and are subject to biases if the assumptions are not met. We found that reasonable violations of the propensity score assumptions could result in estimated absolute differences in 18-month mortality between patients cared for by cardiologist and those treated by generalists ranging from 0.5% to 6%. Similarly, under various assumptions regarding potential violations of the instrumental variable assumptions, causal estimates ranged from a 3.6% absolute decrease in mortality associated with specialty care equal to a 2.2% increase in 18-month mortality. Basing policy decisions on these results requires careful assessment of the assumptions using both observed data and clinical knowledge.

Second, if there is heterogeneity in the impact of cardiology care across strata of patients, propensity score and instrumental variable estimates of causal effects may differ even if the assumptions underlying each are valid. Instrumental variable methods estimate the effect of specialty care on the subset of patients for whom density of cardiologists determined their care. In contrast, propensity score methods based on matched samples estimate the impact of cardiology care in cardiology patients for whom suitable matches can be found.

In Table 6 we report estimated characteristics of the reference population for the causal estimates of specialty care from the propensity score analyses reported in Section 4.1. Estimated specialty effects within strata of estimated propensity scores suggest that there was heterogeneity in the effect of specialty care. In particular, patients least likely to receive specialty care were estimated to have 11.5% lower absolute mortality at 18 months (quintile 1), while the estimated effect in the population of patients typically receiving specialty care was only 2%. Examining the observed characteristics of patients in each of these quintiles suggest that older patients with higher rates of comorbidities benefited most from cardiology care.

Table 6. Characteristics of reference populations for propensity score estimates.

	Quintile			Combined Across Strata		
	1	3	5	Full Sample (equally weighted)	Cardiology Patients (weighted to cardiology)	Matched Sample
Average causal effect (% Diff.)	-11.5	-1.2	1.0	-3.1	-2.2	-3.0
Mean Age	75.9	74.0	71.2	73.8	73.4	74.2
Male (%)	41.7	56.2	70.7	55.0	57.3	52.1
White (%)	87.8	92.2	98.6	92.8	93.7	91.5
COPD (%)	27.1	17.2	6.3	17.3	15.6	19.1
Diabetes (%)	41.3	32.9	21.4	31.8	30.1	33.6
Independently Mobile (%)	70.3	79.3	90.3	80.7	82.3	78.9
Prior MI (%)	27.0	29.8	31.8	28.4	28.7	27.7
Prior CHF (%)	23.7	14.8	9.1	16.1	14.7	17.0
CHF during hospitalization (%)	45.2	39.5	29.6	38.5	37.1	41.1
Recurrent chest pain (%)	22.7	28.9	33.7	28.1	29.0	27.1

COPD = chronic obstructive pulmonary disease, MI = myocardial infarction, CHF = congestive heart failure. Observed characteristics of the patients in the first, third, and fifth quintile of estimated propensity score, in the full sample, in patients observed to receive cardiology care, and in the matched sample. We also report the average causal effect of specialty care estimated within the first, third, and fifth quintile; estimated by combining strata-specific differences using an equally weighted average and an average weighted to the cardiology patients; and estimated in the matched sample

IV methods estimate the effect of specialty care for the subset of patients for whom the instrument determines treatment assignment. In contrast to the propensity score analyses, we cannot directly identify characteristics of this population and must infer them based on clinical knowledge of treatment decisions. Based on the patterns observed in Table 6, we speculate that the subpopulation for whom density of cardiologists in the area determined receipt of cardiology care were younger patients with lower rates of co-morbidities. The propensity-score results suggest that these patients were the least likely to benefit from specialty care and this pattern may explain in part why we found no significant benefit associated with specialty care in the IV analyses.

## 5. Discussion

We illustrated and compared two analytic approaches, propensity scores and an instrumental variable, for estimating the causal effect of outpatient specialty care on mortality in AMI survivors. Using both approaches we found reductions in 18-month mortality associated with ambulatory cardiology care. However, reasonable deviations from the assumptions underlying each method led to estimated differences in mortality ranging

from a 6% absolute reduction in mortality to a 2% absolute increase among patients who received cardiology care.

Choosing an analytic strategy depends on both data availability and the policy question of interest. Propensity-score analyses rely on the assumption that conditional on observed data, treatment is randomly assigned. The plausibility of this assumption depends in part on the extent of observed characteristics available to include in the propensity score model. This study was unusual compared to many in health policy in that we had a large number of detailed clinical variables abstracted from medical records. However, this study lacked information regarding patients' clinical course after discharge, and we could not exclude the possibility that unobserved characteristics of the patients associated with both treatment and mortality resulted in biased estimates. Collection of medical record data in the ambulatory setting could be expensive and complicated if patients receive ambulatory care from multiple physicians, but it would enhance the current study by allowing the control of these potential confounding variables using propensity score methods or extensions of these methods to allow for time-varying confounders and treatment.

In contrast, instrumental variable approaches rely on the identification of good instruments. Candidate instruments must meet five main criteria as outlined in Section 3.2.1. First, IV estimates depend on the ability of the instruments to generate variability in treatment. The greater the ability of the instrument to predict treatment the larger the size of the "compliant" population. Thus IV analyses with highly predictive instruments are less subject to bias (Angrist, Imbens, and Rubin, 1996), estimate effects with greater precision and thus have increased power, and are generalizable to a larger subset of the population. In our study of post-MI ambulatory care, use of specialists ranged from 57% to 72% across the five quintiles of density of cardiologists. This is similar in magnitude to the variability induced by other instruments that have been employed in health policy research. For example, McClellan, McNeil, and Newhouse (1994) observed angiography rates that ranged from 15% to 27% in groups of patients stratified according to the differential distance to a hospital with angiography capabilities. However, we may have been able to increase our precision by employing a more refined measure of cardiology supply, for example, the distance between patient's residence and the nearest outpatient cardiology practice, as an instrument for ambulatory cardiology care, but this information was not available in our study. Second, instrumental variable analyses assume that the instrumental variable is randomly assigned (or that the assignment mechanism is ignorable conditional on observed data). In this example patients differing in density of cardiologists in their county did not differ in many observed clinical characteristics. However, patients living in areas with a low supply of cardiologists were more likely to live in rural areas and differences between urban and rural populations may bias causal estimates. We estimated causal effects within the sample of patients admitted to urban hospitals using IV methods and found similar results to those estimated with the full sample. Third, treatments and outcomes are assumed to be unrelated to the treatment status of other patients. Variability in access to specialists across geographic areas may lead to a violation of this assumption. The impact of such potential violations on causal estimates and estimation techniques that do not rely on this assumption are topics that deserve further study. Fourth, instruments are

assumed to affect outcomes only through their effect on treatment. In this example, there were many observed violations of this assumption and we performed stratified analyses to control for some of the hospital and inpatient treatments that were observed to be correlated with the supply of cardiologists. However, we cannot exclude the possibility that unobserved treatments associated with the supply of cardiologists resulted in biased estimates of the specialty effect. Finally, the instrument is assumed to affect treatment monotonically such that if a patient in a low-density area received cardiology care he or she would also have received cardiology care if he or she lived in a high-density area. In general, it can be very difficult to identify instruments that meet these criteria. In our example, density of cardiologists met the criteria in only one of seven states. An alternative instrument that could be used in the larger dataset would have improved upon the relatively imprecise instrumental variable estimates, however, we were unable to identify such an instrument.

The policy question of interest is also important in choosing an analytic strategy. Observed characteristics of the subset of patients used to compute the propensity-score causal estimates allow us to identify characteristics of the reference population and thus make recommendations for individual patients. However, to estimate the impact of increasing the supply of cardiologists, the IV results are more applicable because they demonstrate the marginal effects of such changes.

Given the difficulty of basing policy decisions on the analysis of observational data, a strategy of employing a variety of methods, with careful assessment of the assumptions underlying each method, may be optimal. For example, our comparative analyses suggest that increasing the supply of cardiologists would result in little improvement in patient outcomes. In contrast, a targeted intervention to increase the use of specialists among patients who are currently least likely to receive cardiology care may substantially improve outcomes in these patients.

### Acknowledgments

Supported by grant RO1-HS09718 from the Agency for Healthcare Research and Quality, Rockville, MD. We gratefully acknowledge Barbara McNeil and Joseph Newhouse, Harvard Medical School, for comments on early versions of the manuscript; two anonymous referees for helpful comments; and Peter Gaccione and Margaret Volya, Harvard University, for programming support.

### References

- J. D. Angrist, G. W. Imbens and D. B. Rubin, "Identification of causal effects using instrumental variables," *Journal of the American Statistical Association*, 434, pp. 444–454, 1996.
- J. D. Angrist and G. W. Imbens, "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American Statistical Association*, 90, pp. 431–442, 1995.



- J. Z. Ayanian, E. Guadagnoli, B. J. McNeil and P. D. Cleary, "Treatment and outcomes of acute myocardial infarction among patients of cardiologists and generalist physicians," *Archives of Internal Medicine*, 1997.
- S. J. Borowsky, R. L. Kravitz and M. Laouri, et al., "Effect of physician specialty on use of necessary coronary angiography," *Journal of the American College of Cardiology*, 26, pp. 1484–1491, 1995.
- A. F. Connors, T. Speroff and N. V. Dawson, et al., "The effectiveness of right heart catheterization in the initial care of critically ill patients," *Journal of the American Medical Association*, 276, pp. 889–897, 1996.
- R. B. D'Agostino, "Tutorial in biostatistics: propensity score methods for bias reduction in the comparisons of a treatment to a non-randomized control group," *Statistics in Medicine*, 17, pp. 2265–2281, 1998.
- M. E. Glickman and S. T. Normand, "The derivation of a latent threshold instrumental variables model," *Statistica Sinica*, 10, pp. 517–544, 2000.
- C. D. Frances, M. G. Shlipak, H. Noguchi, P. Heidenreich and M. McClellan, "Does physician specialty affect the survival of elderly patients with myocardial infarction?" *Health Services Research*, 35, pp. 1093–1116, 2000.
- K. M. Harris and D. K. Remler, "Who is the marginal patient? understanding instrumental variables estimates of treatment effects," *Health Services Research*, 33, pp. 1337–1360, 1998.
- K. Hirano, G. W. Imbens, D. B. Rubin and X. H. Zhou, "Assessing the effect of an influenza vaccine in an encouragement design," *Biostatistics*, 1, pp. 69–88, 2000.
- P. Holland, "Statistics and causal inference," *Journal of the American Statistical Association*, 81, pp. 945–970, 1986.
- G. W. Imbens and J. D. Angrist, "Identification and estimation of local average treatment effects," *Econometrica*, 62, pp. 467–475, 1994.
- J. Johnston and J. DiNardo. *Econometric Methods*, McGraw-Hill: New York, 1997.
- J. G. Jollis, E. R. DeLong and E. D. Peterson, et al., "Outcome of acute myocardial infarction according to the specialty of the admitting physician," *New England Journal of Medicine*, 335, pp. 1880–1887, 1996.
- T. A. Marciniak, E. F. Ellerbeck and M. J. Radford, et al., "Improving the quality of care for medicare patients with acute myocardial infarction: results from the cooperative cardiovascular project," *Journal of the American Medical Association*, 279, pp. 1351–1359, 1998.
- M. McClellan, B. J. McNeil and J. P. Newhouse, "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?" *Journal of the American Medical Association*, 272, pp. 859–866, 1994.
- I. S. Nash, D. B. Nash, V. Fuster, "Do cardiologists do it better?" *Journal of the American College of Cardiology*, 29, pp. 274–278, 1997.
- C. A. Polanczyk, L. E. Rohde, L. Goldman, E. F. Cook, E. J. Thomas, E. R. Marcantonio, C. M. Mangione and T. H. Lee, "Right heart catheterization and cardiac complications in patients undergoing noncardiac surgery: an observational study," *Journal of the American Medical Association*, 286, pp. 309–314, 2001.
- P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, pp. 41–55, 1983a.
- P. R. Rosenbaum and D. B. Rubin, "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome," *Journal of the Royal Statistical Society, Series B*, 45, pp. 212–218, 1983b.
- P. R. Rosenbaum and D. B. Rubin, "Reducing bias in observational studies using subclassification on the propensity score," *Journal of the American Statistical Association*, 79, pp. 516–524, 1984.
- P. R. Rosenbaum and D. B. Rubin, "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score," *American Statistician*, 39, pp. 33–38, 1985a.
- P. R. Rosenbaum and D. B. Rubin, "The bias due to incomplete matching," *Biometrics*, 41, pp. 103–166, 1985b.
- D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66, pp. 688–701, 1974.
- D. B. Rubin, "Bayesian inference for causal effects," *The Annals of Statistics*, 6, pp. 34–58, 1978.
- D. B. Rubin, "Using Multivariate Matched Sampling and Regression Adjustment to control bias in observational studies," *Journal of the American Statistical Association*, 74, pp. 318–328, 1979.
- D. R. Thieman, J. Coresh, W. J. Oetgen and N. R. Powe, "The association between hospital volume and survival after acute myocardial infarction in elderly patients," *New England Journal of Medicine*, 21, pp. 1640–1648, 1999.