Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods

JAMA. 2007 January 17; 297(3): 278–285.

Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods

Thérèse A. Stukel, PhD, **Elliott S Fisher, MD, MPH**, **David E. Wennberg, MD, MPH**, **David A. Alter, MD, PhD**, **Daniel J. Gottlieb, MS**, and **Marian J. Vermeulen, MHSc**
Institute for Clinical Evaluative Sciences, Toronto, Ontario (Drs Stukel and Alter, and Ms Vermeulen); Center for the Evaluative Clinical Sciences, Dartmouth Medical School, Hanover, NH (Drs Stukel and Fisher, and Mr Gottlieb); Department of Health Policy, Management, and Evaluation, University of Toronto, and Clinical Epidemiology and Health Care Research Program, Sunnybrook Health Sciences Centre, Toronto, Ontario (Drs Stukel and Alter); Veterans Administration Outcomes Group, White River Junction, Vt (Dr Fisher); Center for Outcomes Research and Evaluation, Maine Medical Center, Portland (Dr Wennberg); and Division of Cardiology and the Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto Rehabilitation institute, and Department of Medicine, University of Toronto, Toronto, Ontario (Dr Alter).

Pokket Sirisreetreerux

Introduction

- In the face of the financial, practical, and ethical challenges inherent in undertaking randomized clinical trials (RCTs), investigators often use observational data to compare the outcomes of different therapies.
- These comparisons may be biased due to prognostically important baseline differences among patients, often as a result of unobserved treatment selection biases.
- Unmeasurable clinical and social interactions in the diagnostic-treatment pathway, and physicians’ knowledge of unmeasured prognostic variables, may affect treatment decisions and outcomes.

Introduction

- Physicians are frequently risk averse in case selection, performing interventions on lower-risk patients despite greater clinical benefit to higher-risk patients.
- Due to the complexity and cost of performing RCTs, there is interest in using observational studies to guide policy statements and clinical protocols, and in generalizing results to the community.

Introduction

- More than 280 000 US Medicare enrollees are admitted to the hospital with acute myocardial infarction (AMI) annually.
- Much of the effort to reduce high mortality rates has focused on invasive diagnostic and therapeutic interventions, such as cardiac catheterization followed by revascularization.
- Recent systematic reviews of RCTs assessing routine invasive vs conservative therapies found between 8% and 21% improved relative survival in the more invasively-treated group.
- A recent population-based observational study found little benefit to invasive therapy in US regions in which medical management was of higher quality.

Introduction

- The author reanalyzed these data to demonstrate how the estimated benefit from invasive therapy depends on the statistical method used to adjust for overt (measured) and hidden (unmeasured) bias.
- Methods included multivariable model risk adjustment, propensity score risk adjustment, and propensity-based matching, which control for overt bias, and instrumental variable analysis, which is a method designed to control for hidden bias as well.

Study Cohort and Data Sources

- They derived the study cohort from the Cooperative Cardiovascular Project, a US national sample of Medicare enrollees hospitalized with first admission for AMI in nonfederal acute care hospitals in 1994–1995.
- The Cooperative Cardiovascular Project comprised clinical data abstracted from medical records during admission, including presentation characteristics, comorbidities, and inpatient treatments, follow up patients for 7 years.
- They restricted analyses to patients eligible for cardiac catheterization with American College of Cardiology/American Heart Association class I (ideal) or class II (uncertain) indications.

Treatment Variables

- We examined whether invasive cardiac treatment predicted long-term mortality.
- Patient-level treatment
  - Patients who receive invasive cardiac treatment are generally younger, healthier, have lower AMI severity, and may differ in unobserved ways from those who do not.
- In contrast, mean AMI admission severity tends to be similar across areas.
- Regional treatment intensity
  - defined as the percentage of eligible patients receiving cardiac catheterization within 30 days of admission for 566 coronary angiography service areas.
- Patients were assigned to the cardiac catheterization rate of their region of residence.

Main Outcome Measure

- Patients were followed up from date of AMI admission (index event) through December 31, 2001.
- The main outcome measure was long-term mortality over 7 years of follow-up.
- Date of death was obtained from the Medicare Denominator file.

Statistical Methods

- All models used the patient as the unit of analysis.
- They developed an AMI severity index using Cox proportional hazards regression models to predict 1-year mortality using all baseline patient characteristics of age, sex, race, socioeconomic status, comorbidities, and clinical presentation (*c* statistic = 0.77).

Statistical Methods

- Cox proportional hazards regression models were used to compare mortality rates between treatment groups, adjusting for 65 patient, hospital, and ZIP code characteristics associated with post-AMI mortality.
- Because patients admitted to the same hospital may have correlated outcomes, survival models incorporated clustering by hospital to adjust the SEs.
- Model fit and proportionality of hazards were assessed using residual analyses.
- Analyses were performed by using the STATA procedure STCOX.

Multivariable Model Risk Adjustment

- The multivariable model risk adjustment model is the conventional modeling approach that incorporates all known confounders, including interactions, into the model.
- Controlling for these covariates produces a risk-adjusted treatment effect and removes overt bias due to these factors.
- Cox proportional hazards regression models were used to compare mortality rates between those patients who did or did not receive cardiac catheterization, adjusted for all 65 covariates.

Propensity Score Risk Adjustment

- The propensity score is the probability of receiving treatment for a patient with specific prognostic factors.
- Within propensity score strata, covariates in treated and control groups are similarly distributed, so that stratifying on propensity score strata removes more than 90% of the overt bias due to the covariates used to estimate the score.
- Propensity scores cannot remove hidden biases except to the extent that unmeasured prognostic variables are correlated with the measured covariates used to compute the score.

Propensity Score Risk Adjustment

- They computed the propensity score by using logistic regression with the dependent variable being receipt of cardiac catheterization, and the independent variables (covariates) being the 65 patient, hospital, and ZIP code variables.
- To provide optimal control for confounding, they computed a second propensity score based on the above covariates and all 3-way interactions of age, sex, race, and these variables (750 variables).
- Propensity scores were categorized into deciles.
- Cox proportional hazards regression models were used to compare mortality rates between those patients who did or did not receive cardiac catheterization, adjusting for propensity decile.

Propensity-Based Matching

- Propensity-based matching is used to select control patients who are similar to patients receiving treatment with respect to propensity score and other covariates, discarding unmatched individuals, thereby matching on many confounders simultaneously.
- Patients receiving cardiac catheterization were matched to the closest control whose propensity score differed by less than 0.10 among those patients within 5 years of age.
- Cox proportional hazard regression models were used to compare adjusted mortality rates between those patients who did or did not receive cardiac catheterization,conditional on matched pair.

Instrumental Variable Analysis

- Instrumental variable analysis is an econometric method used to remove the effects of hidden bias in observational studies.
- An instrumental variable has 2 key characteristics:
  - it is highly correlated with treatment and does not independently affect the outcome
  - it is not associated with measured or unmeasured patient health status.

Instrumental Variable Analysis

- We demonstrate that regional cardiac catheterization rate can serve as an effective instrumental variable because prognostic factors related to mortality, such as mean AMI severity, are similar across regions that have dramatically different cardiac catheterization rates.
- The instrumental variable behaves like a natural randomization of patients to regional “treatment groups” that differ in likelihood of receiving cardiac catheterization.
- Rather than compare patients with respect to the actual treatment received since this might be biased, instrumental variable analysis compares groups of patients that differ in likelihood of receiving cardiac catheterization.

Instrumental Variable Analysis

- Instrumental variable models produce adjusted estimates of treatment effect on mortality at one time point, on an absolute rather than a relative scale.
- We first estimated adjusted absolute mortality differences 1 and 4 years after index admission between patients receiving vs not receiving cardiac catheterization, using multiple linear regression with the dependent variable being mortality considered as a binary variable.
- We then estimated instrumental variable-adjusted mortality differences, with the instrumental variable being the regional cardiac catheterization rate, using the STATA procedure IVREG.
- All models controlled for all 65 covariates.
- Cox proportional hazards regression models were used to estimate relative mortality rates across quintiles of regional cardiac catheterization rate, demonstrating an implicit use of the instrumental variable technique

Results

Standard Risk-Adjustment

- The study cohort consisted of 122,124 patients, 73 238 (60%) of whom received cardiac catheterization within 30 days.
- Patients who received cardiac catheterization were younger, men, had lower AMI severity, and were more likely to be admitted to high volume hospitals.
- Mean cardiac catheterization propensity scores ranged from 0.16 to 0.90 across propensity deciles, with excellent discrimination between treatment groups (*c* statistic = 0.76).

Standard Risk-Adjustment

- The distribution of key confounders, such as predicted 1-year mortality, age, arid history of congestive heart failure, was similar within propensity deciles for those patients with and without cardiac catheterization, except possibly in the lowest decile. (TABLE 2).

Standard Risk-Adjustment

- Propensity-based matching produced 31,193 matched pairs with standardized differences in patient characteristics of less than 10%, indicating a high degree of similarity in the distributions of prognostic variables (Table 1).
- No match was found for 42 045 patients receiving cardiac catheterization.
- Cardiac catheterization was associated with an approximate 50% relative decrease in mortality rate, using multivariable model risk adjustment, propensity score risk adjustment, or propensity-based matching (TABLE 3).

Table 1

Select Baseline Characteristics According to Receipt of Cardiac Catheterization*

	Overall Cohort			Propensity-Based Matched Cohort			
	Received Cardiac Catheterization Within 30 Days			Received Cardiac Catheterization Within 30 Days			
	No (n=48 886)	Yes (n=73 238)	Standardized Difference	No (n=31 193)	Yes (n =31 193)	Standardized Difference	Unmatched Patients Receiving Cardiac Catheterization (n=42 045)
Predicted 1 -year mortality (AMI severity), mean (SD) †	32.3 (18.3)	20.9 (13.3)	73.7	26.8 (15.5)	27.8 (12.5)	6.3	15.8 (7.5)
Demographics							
Age range, y							
65–74	40.2	64.4	49.9	45.2	45.3	0.1	78.6
75–84	59.8	35.6	49.9	54.8	54.7	0.1	21.4
Men	49.7	58.4	17.6	53.2	49.6	7.2	65.0
Black	7.5	4.8	11.3	5.7	6.6	3.7	3.5
Social Security income ≥ $2600	30.0	29.7	0.9	30.2	30.2	0.1	29.2
Comorbidities							
History of angina	44.1	49.9	11.8	46.0	45.6	0.9	53.2
Previous myocardial infarction	32.9	26.4	14.3	28.7	31.9	6.8	22.3
Previous revascularization	17.8	20.9	7.7	18.0	20.2	5.7	21.3
Congestive heart failure	27.2	10.4	45.7	16.6	14.3	4.4	4.6
Diabetes mellitus	36.6	28.6	17.1	31.8	34.1	4.9	24.5
Peripheral vascular disease	12.8	9.9	12.0	10.6	11.5	2.8	7.8
Chronic obstructive pulmonary disease	24.9	17.6	18.3	20.9	23.3	5.9	13.3
Smoker‡	16.1	18.0	5.0	16.5	17.0	1.2	18.8
AMI clinical presentation characteristics							
Non-ST-segment elevation AMI	41.8	38.9	5.9	39.8	40.1	0.8	38.0
Shock	1.9	1.5	3.0	1.8	2.3	3.4	0.9
Hypotension	3.5	2.3	7.4	3.1	3.6	2.6	1.2
Received CPR	1.8	1.6	1.6	2.3	3.5	7.3	0.2

Table1

	Overall Cohort			Propensity-Based Matched Cohort			
	Received Cardiac Catheterization Within 30 Days			Received Cardiac Catheterization Within 30 Days			
	No (n=48 886)	Yes (n=73 238)	Standardized Difference	No (n=31 193)	Yes (n =31 193)	Standardized Difference	Unmatched Patients Receiving Cardiac Catheterization (n=42 045)
Peak creatinine kinase >1000 U/L	29.1	32.4	7.2	31.7	31.8	0.2	32.9
Hospital characteristics							
Annual AMI volume >200 patients	20.1	30.4	23.6	22.9	20.5	5.6	37.8
Mortality§							
Died within 1 y	38.6	14.2		34.6	19.0		10.6
Died within 4 y	62.0	27.8		55.4	36.3		21.4

Table2

Distribution of Select Covariates by Propensity Score Deciles, According to Receipt of Cardiac Catheterization

	Decile (Range) of Propensity Score*									
	1 (0.00–0.26)	2 (0.26–0.40)	3 (0.40–0.50)	4 (0.50–0.58)	5 (0.58–0.65	6 (0.65–0.70)	7 (0.70–0.75)	8 (0.75–0.80)	9 (0.80–0.85)	10 (0.85–0.98)
No. of patients										
No cardiac catheterization	10021	8219	6873	5763	4834	3997	3283	2628	2060	1208
Cardiac catheterization	2191	3993	5340	6449	7378	8215	8930	9585	10151	11006
Predicted 1 -year mortality, %†										
No cardiac catheterization	54.5	39.2	31.8	27.5	23.4	20.0	17.3	15.3	14.0	13.6
Cardiac catheterization	51.2	38.9	31.8	27.4	23.5	20.0	17.3	15.3	13.5	12.8
Mean age, y‡										
No cardiac catheterization	79.4	78.0	77.0	75.5	74.3	72.9	71.9	70.8	70.1	70.0
Cardiac catheterization	79.3	77.9	76.8	75.7	74.3	73.0	71.8	70.9	70.0	69.9
History of congestive heart failure, %										
No cardiac catheterization	59.8	40.0	27.0	18.8	10.8	7.3	4.2	2.7	2.0	2.1
Cardiac catheterization	61.4	40.0	26.5	16.7	10.5	5.7	3.6	2.5	2.0	1.7

*Propensity scores were rounded to 2 decimal points. There was no overlap across deciles,

†Predicted 1 -year mortality was computed using the Cox proportional hazards regression mode!, including all baseline patient characteristics of age, sex, race, socioeconomic status, comorbidities, and clinical presentation.

‡SD for age was 4.3 years.

Table 3

Adjusted Relative Mortality Rate Associated With Receipt of Cardiac Catheterization Among Patients With AMI Using Standard Risk-Adjustment Methods

Risk-Adjustment Method	Relative Mortality Rate (95% CI)
Unadjusted survival model	0.364 (0.358–0.370)
Multivariable survival model (65 covariates)	0.510(0.502–0.519)
Survival models using simple propensity score*	
Propensity deciles alone	0.538 (0.529–0.547)
Propensity deciles plus all covariates	0.520 (0.511–0.529)
Survival models using complex propensity score†	
Propensity deciles alone	0.540 (0.531–0.549)
Propensity deciles plus all covariates	0.522 (0.513–0.531)
Survival models using propensity-based matching cohort	
Match within ±0.05 of propensity score and 5 y of age	0.538 (0.518–0.558)
Match within ±0.10 of propensity score and 5 y of age	0.528 (0.514–0.542)
Match within ±0.15 of propensity score and 5 y of age	0.511 (0.499–0.523)

Abbreviations: AMI, acute myocardial infarction; CI, confidence interval.

*Simple propensity score included all 65 patient, hospital, and ZIP code characteristics.

†Complex propensity score included all patient, hospital, and ZIP code characteristics and all interactions of age, sex, and race with the other characteristics (750 variables).

Instrumental Variable Analyses

- Mean cardiac catheterization rate within 30 days ranged from 29% to 82% across regions and 43% to 65% across cardiac catheterization quintiles.
- Table 4 reports selected baseline characteristics of study patients, according to quintiles of regional cardiac catheterization rate.
- Although there were small differences in specific risk factors, mean predicted 1-year mortality, our summary measure, of AMI severity, was remarkably similar across regions (quintile 1 [lowest], 26.1%; quintile 2, 26.0%; quintile 3, 25.5%; quintile 4, 25.3%; and quintile 5 [highest], 24.6%).

Instrumental Variable Analyses

- The balance in the distribution of all measured risk factors across regions provides reasonable evidence to infer that the distribution of unmeasured risk factors is likely balanced across regions as well.
- The wide range of cardiac catheterization rates and the similarity in average patient characteristics lend support to regional cardiac catheterization rates being a strong, valid instrumental variable.

Table4

Table 4

Selected Baseline Characteristics and Adjusted Relative Mortality Rates Across Quintiles of Regional Cardiac Catheterization Rate

	Quintile (Range) of Regional Cardiac Catheterization Rate, %				
	1 (29.2–48.1)	2 (48.2–53.0)	3 (53.1–56.3)	4 (56.4–60.2)	5 (60.3–82.3)
No. of patients	24872	24184	24718	24063	24287
Cardiac catheterization rate	42.8	50.6	54.7	58.0	65.0
Mean predicted 1 -year mortality (AMI severity)*	26.1	26.0	25.5	25.3	24.6
Demographics Age range, y 65–74	53.3	54.4	54.6	55.6	55.6
75–84	46.7	45.6	45.4	44.4	44.4
Men	53.7	54.2	55.0	55.6	56.4
Black	4.1	8.1	6.3	5.5	5.4
Social Security income ≥ $2600	30.4	28.2	33.4	27.9	29.1
Comorbidities History of angina	50.1	48.3	47.8	47.6	44.0
Previous myocardial infarction	30.1	29.8	29.2	28.7	26.9
Previous revascularization	16.5	18.6	20.8	20.2	22.1
Congestive heart failure	18.4	18.0	17.3	16.9	15.1
Diabetes mellitus	32.9	32.5	32.3	31.3	30.0
Peripheral vascular disease	10.5	10.9	11.0	10.4	10.0
Chronic obstructive pulmonary disease	21.1	20.2	20.3	20.3	20.7
Smoker†	16.7	16.7	17.0	18.0	17.9
AMI clinical presentation characteristics Non–ST-segment elevation AMI	40.4	41.2	40.5	39.3	39.0
Shock	1.6	1.6	1.6	1.7	1.7
Hypotension	2.8	2.9	2.6	2.8	2.7
Received CPR	1.6	1.7	1.7	1.8	1.7
Peak creatinine kinase >1000 U/L	30.3	30.5	30.4	31.7	32.6
Hospital characteristics Annual AMI volume >200 patients	24.2	24.6	30.4	28.5	23.8
Mortality‡ Died within 1 y	25.0	24.8	23.9	23.7	22.3

Instrumental Variable Analyses

- Unadjusted 4-year mortality was 33.9% points lower in patients receiving cardiac catheterization vs patients not receiving cardiac catheterization (Table 5).
- Adjusted differences were attenuated, and instrumental variable estimates were further attenuated, producing an instrumental variable–adjusted absolute mortality decrease of 9.7% points.
- This corresponds with an approximate instrumental variable–adjusted relative mortality rate of 0.84 (95% confidence interval [CI], 0.79–0.90).
- Similar patterns were found at 1 year.
- The relative mortality rate in regions with the highest (>60.2%) compared with the lowest (<48.2%) cardiac catheterization rates was 0.95 (95% CI, 0.92–0.97), demonstrating an implicit use of instrumental variable techniques (Table 4).

Table 5

Adjusted Mortality Differences Associated With Cardiac Catheterization Among Patients With AMI Using Linear Regression and Instrumental Variable Methods

Risk-Adjustment Method	Absolute Mortality Difference (Δ) (SE)	Adjusted Relative Mortality Rate (95% CI)[*]
1 -Year mortality		
Unadjusted	−0.244 (0.002)	0.37 (0.35–0.38)
Multiple linear regression[†]	−0.162 (0.002)	0.58 (0.57–0.59)
Instrumental variable, adjusted[‡]	−0.054 (0.015)	0.86 (0.78–0.94)
4-Year mortality		
Unadjusted	−0.339 (0.003)	0.45 (0.44–0.46)
Multiple linear regression[†]	−0.207 (0.003)	0.67 (0.66–0.68)
Instrumental variable, adjusted[‡]	−0.097 (0.016)	0.84 (0.79–0.90)

Abbreviations: AMI, acute myocardial infarction; CI, confidence interval.

[*] Adjusted relative mortality rate is approximately $1 + \Delta/m_{\text{noCATH}}$, where Δ is the adjusted absolute mortality difference between patients with and without cardiac catheterization, and m_{noCATH} is the Kaplan-Meier mortality rate among those patients without cardiac catheterization.

[†] Linear regression of mortality (binary variable) against all 65 observed patient, hospital, and ZIP code characteristics.

[‡] Instrumental variable analysis using mortality (binary variable) as the dependent variable and instrumental variable as regional cardiac catheterization rate for the 566 coronary angiography service areas, adjusted for all 65 observed patient, hospital, and ZIP code characteristics.

Discussion

- Within a large observational data set, the estimated association of invasive cardiac treatment with long-term mortality is sensitive to the analytic method used.
- Cardiac catheterization predicted a 50% relative decrease in mortality using standard risk-adjustment methods, including a rigorous propensity-based matching analysis, even after accounting for a clinically rich set of prognostic variables.
- Using instrumental variable methods, the associated relative decrease in mortality was approximately 16%. When estimated treatment associations vary 3-fold depending on the method used, several questions should come to mind.
- Do the results have face validity? The survival benefits of routine invasive care from RCTs are between 8% and 21%.
- Results in RCTs are optimized and tend to overestimate the relative benefits achievable in routine clinical practice.

Discussion

- The overestimate of benefit using standard modeling is likely due to residual confounding related to the selection of lower-risk patients for cardiac catheterization.
- The magnitude of bias may be greater than usual because receiving catheterization required surviving from admission until this treatment.
- Even controlling for complete information on patients’ admission severity could not eliminate this important survival bias.
- Such situations are not unusual in observational studies of surgical procedures.

Discussion

- The instrumental variable estimate of a 16% relative survival benefit was closer to RCT results because we used a strong, valid instrumental variable.
- Although there may be residual unmeasured regional illness differences, this is unlikely since predicted mortality was estimated using strongly prognostic risk factors and was similar for measured covariates across regions.
- Instruments that are more predictive of treatment produce less biased estimates and smaller SEs, and provide closer approximations to the average population effects from RCTs.

When are standard statistical methods likely to produce unbiased findings?

- The distribution of unmeasured prognostic factors are more likely to be similar when considering therapies with similar clinical indications and risk.
- Randomized clinical trials and observational studies show the greatest similarities under such conditions.
- Observational studies of invasive procedures are more prone to bias because patients who are candidates for surgery often differ in unmeasurable ways from patients who are not.

Is the similarity between multivariable and propensity model estimates expected?

- The utility of instrumental variable analyses depends on finding a strong, valid instrumental variable and careful interpretation.
- The instrumental variable estimate measures the treatment effect on the “marginal” population. This excludes those patients who would “always” or “never” receive cardiac catheterization, focusing on patients with uncertain indications whose likelihood of being treated depends on local clinical judgment and catheterization laboratory supply.

When are nontraditional approaches useful?

- Instrumental variable analyses are most, suited to inform policy decisions.
- Because region or physician is often the level at which policy and resource allocation decisions are made, such studies assess the effects of health system factors on patient outcomes.
- These studies answer policy-relevant questions, such as “What are the benefits of increasing the regional cardiac catheterization laboratory capacity?”, because this would increase the routine provision of invasive services to the AMI population.

Conclusion

- Randomized clinical trials cannot be undertaken in all situations in which evidence is needed to guide care.
- Well-designed observational studies are still needed to assess population effectiveness and to extend results to a general population setting.
- This study serves as a cautionary note regarding their analysis and interpretation.
    - First, propensity scores and propensity-based matching have the same limitations as multivariable risk adjustment model methods, arid are no more likely to remove bias due to unmeasured confounding when strong selection bias exists.
    - Second, instrumental variable analyses may remove both overt and hidden biases but are more suited to answer policy questions than to provide insight into a specific clinical question for a specific patient.

Conclusion

- Caution is advised regarding clinical protocols and policy statements for invasive care based on expected mortality benefits derived from traditional multivariable modeling and propensity score risk adjustment of observational studies

Thank you