OBSERVATIONAL METHODS IN CARDIOVASCULAR OUTCOMES RESEARCH

Jerome Jeffrey Federspiel

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Health Policy and Management.

Chapel Hill
2013

Approved by:

Morris Weinberger, Ph.D.

Patricia P. Chang, M.D., M.H.S.

Donna A. Gilleskie, Ph.D.

G. Mark Holmes, Ph.D.

Sally C. Stearns, Ph.D.

To my infinitely patient wife, Mary Clay.

ABSTRACT

JEROME JEFFREY FEDERSPIEL: Observational Methods in Cardiovascular Outcomes Research
(Under the direction of Morris Weinberger)

Compared to randomized trials, observational methods have several advantages for cardiovascular outcomes research, including lower cost, reduced risk to subjects, and improved generalizability. Here, I leverage recently-developed observational methodologies to address questions that would be difficult or impossible to conduct using randomized trials. Study 1 centers on patterns and predictors of stress test use after elective percutaneous coronary intervention (PCI). I find that: (1) stress testing is commonly performed after elective PCI in a pattern suggestive of routine testing; (2) risk factors thought to increase the potential value of such testing are paradoxically associated with lower use of testing, and (3) the rate of stress testing use varies strongly across the facilities participating in the national CathPCI Registry in a manner that is associated with higher rates of repeat revascularization procedures without reduction in death or myocardial infarction. Study 2 centers on the imaging modality (echocardiography versus nuclear imaging) chosen for patients receiving an exercise stress test after PCI. While many comparisons of test performance (*i.e.*, sensitivity and specificity) have been made for echocardiography and nuclear imaging, little is known about the implications of test choice on clinical outcomes or resource use. I find that patients who receive echocardiography received fewer subsequent coronary catheterization or revascularization procedures, but more repeat stress tests, than do nuclear testing patients. No differences in rates of death or readmission for myocardial infarction

were noted. Study 3 illustrates the use of newly-developed instrumental variables methodologies for outcomes research. While conventional instrumental variables techniques are only able to estimate a local average treatment effect, or the effect of a treatment on an unidentifiable "marginal" population of patients, newer methodologies allow for the estimation of more relevant estimands, such as the average treatment effect or effect of treatment on those patients receiving the treatment in clinical practice. We evaluated the effectiveness of drug-eluting versus bare metal coronary stents using these new methods, finding evidence that drug-eluting stents are safe and effective in patients receiving them, but that there is considerable heterogeneity in treatment response.

ACKNOWLEDGEMENTS

I sincerely appreciate my committee's flexibility and patience with me as I worked through a rather unconventional dissertation, both in format and in execution. Particular thanks are owed to Morris Weinberger, for his steady guidance in bringing this dissertation to completion, and to Sally Stearns, who has supported and guided my work for the past four years.

In addition to my committee, I have benefitted from the support and guidance of numerous individuals outside UNC. Special thanks are due to a number of faculty members at the Duke Clinical Research Institute, including Pamela Douglas, Patricia Cowper, Cynthia Green, Bimal Shah, and Kevin Anstrom. Additional guidance was provided by Frederick Masoudi (University of Colorado) and Leslee Shaw (Emory University).

I am grateful to the members of my dissertation writing group (Alice Fortune-Greeley, Kimberly Geissler, Mona Kilany, Ashley Kranz, Elizabeth La, Laurel Trantham) for their support over the years. Completing a dissertation is a unique experience which cannot be appreciated unless experienced and I am grateful for the opportunity to share this experience (both the rewards and challenges) with a wonderful group of individuals.

Finally, I greatly appreciate the financial support of the American people and the people of North Carolina, without which this dissertation and indeed my entire joint degree training would not have been possible. I was funded by a Ruth L. Kirschstein National Research Service Award for Individual Predoctoral MD/PhD Fellows (F30-HL110483) from

the National Heart, Lung, and Blood Institute while completing this dissertation. Earlier phases of my medical training were funded by an award to the UNC MD/PhD Program (T32-GM008719) and funding from the State of North Carolina. Data for Study 1 was funded under contract no. HHSA290200500321 Task Order No. 4 from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services as part of the Developing Evidence to Inform Decisions about Effectiveness (DEcIDE) program. Furthermore, the data in Study 1 was provided courtesy of the National Cardiovascular Data Registry (NCDR). The views expressed in this manuscript represent those of the authors, and do not necessarily represent the official views of the NCDR or its associated professional societies identified at www.ncdr.com. Data for Studies 2 and 3 were funded by the National Institutes on Aging (R01-AG025801). All options are the responsibility of the author; statements in this dissertation should not be construed as being endorsed by the U.S. Department of Health and Human Services.

PREFACE

This dissertation is structured in an unconventional manner to better reflect the close topical unity of Papers 1 and 2 versus Paper 3, as well as differences between the level of detail permitted in most biomedical publications and the level of detail expected in a doctoral dissertation. As a result, while a brief general introduction is provided, Studies 1 and 2 are also preceded by an additional introductory chapter providing additional context on cardiovascular imaging.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

ACC – American College of Cardiology

ACCF – American College of Cardiology Foundation

ACS – Acute coronary syndrome

AHA – American Heart Association

AHRQ – Agency for Healthcare Research and Quality

ATE – Average treatment effect

AUC – Appropriate use criteria

BMS – Bare metal stent

CABG – Coronary artery bypass grafting

CER – Comparative effectiveness research

CI – Confidence interval

CMS – Centers for Medicare and Medicaid Services

CVD – Cardiovascular disease

DES – Drug-eluting stent

ECG – Electrocardiogram

Echo – Echocardiography

ESC – European Society for Cardiology

FFS – Fee for service

HR – Hazard ratio

HCPCS – Healthcare Common Procedure Coding System

ICD-9-CM – International Classification of Diseases, Ninth Edition, Clinical Modification

IQR – Interquartile range

IV – Instrumental variable

LATE – Local average treatment effect

LIV – Local instrumental variable

MedPAR – Medicare Provider Analysis and Review

MI – Myocardial infarction

MTE – Marginal treatment effect

NCDR – National Cardiovascular Data Registry

NIH – National Institutes of Health

NSTEMI – Non ST-elevation myocardial infarction

OR – Odds ratio

PeT – Person-centered treatment effects

PCI – Percutaneous coronary intervention

STEMI – ST-elevation myocardial infarction

TT – Effect of treatment on the treated

TUT – Effect of treatment on the untreated

UA – Unstable Angina

US – United States

ZIP – Zone Improvement Plan

1. INTRODUCTION

In economically-developed countries such as the United States, the burden of cardiovascular disease (CVD) on morbidity and mortality has steadily declined for over fifty years. Between 1968 and 2006, the age-adjusted CVD mortality rate in the US fell by 65 percent (1). While lifestyle changes such as reduced smoking rates and improved nutrition are responsible for some of this decline, half of the reduction in death due to coronary heart disease that occurred in the 1980s and 1990s was due to improved medical treatment (2-4).

Improvements in treatment are certainly due in part to the development of novel treatments for CVD, but health services research's role in optimizing the delivery of cardiovascular care is also a critical driver of improved CVD outcome (5-8). For example, despite its low cost and compelling evidence of survival benefit, aspirin was prescribed for only 60 percent of Medicare beneficiaries who experienced a myocardial infarction in 1992-1993 (9). After health services researchers brought attention to this quality issue, prescription rates increased to 90% by 2001 (10). Similarly, the development of percutaneous coronary intervention (PCI) to terminate certain types of myocardial infarction (ST-elevation myocardial infarction) was a lifesaving clinical breakthrough (11). However, recent efforts by health services researchers have expanded the value of this lifesaving, but time-critical, procedure by developing strategies to rapidly transport patients to facilities able to provide the procedure and reducing the intra-facility delay between entering a facility and receiving PCI (12, 13). Simply put, cardiovascular health services research improves quality of care and has a long tradition of improving patients' lives.

Despite impressive progress, CVD remains an important clinical and public health problem. It continues to be the most common cause of mortality in the United States, responsible for 34 percent of deaths in 2006 (1). By 2030, it is projected that 41 percent of the United States population will have been diagnosed with some form of CVD (14). Additionally, improvements in outcome have come at growing financial cost to both patients and third party payers. Direct medical expenditures on CVD were estimated to exceed $324 billion in 2010, and the growth in CVD-related medical care costs exceeds even the general rate of medical care inflation (15). As a result, CVD-related care is an increasingly important component of the overall explosive growth in American medical care spending, which ultimately threatens the ability of patients to access appropriate medical care (16). Thus, the agenda for CVD-related health services research in the 21st century must include identifying approaches that continue the trend of steadily improving patient outcomes in a manner that is sensitive to resource use.

Randomized trials are considered the gold standard methodology to evaluate the efficacy of interventions, because a well-conducted randomized trial is the methodology that is least vulnerable to bias due to confounding (*i.e.*, treatment assignment is exogenous by construction). However, trials are conducted under highly controlled conditions, with a narrow range of patients, which limits their generalizability to real world patients and practices. The use of observational data has several compelling advantages for CVD-related health services research (17):

1. Observational data allow for the evaluation of treatment effects that reflect utilization patterns in the "real world" of clinical practice, rather than the selected populations studied in trials. In addition to clinical exclusions that systematically reduce the number of typical

patients with comorbid conditions, current randomized trial design and recruitment efforts result in cohorts that under-represent the elderly, women, and racial or ethnic minorities (18). Furthermore, because patients must consent to participate in trials, those who agree to study enrollment differ in systematic ways from patients who refuse to consent (19, 20). Because the effects of most medical treatments are likely heterogeneous (*i.e.*, treatments work better in some patients than others), the fact that patients enrolled in clinical trials may be systematically different than the typical patient eligible for treatment is a significant limitation of randomized trials.

2. The sample sizes attainable with some observational datasets allows for the assessment of rare outcomes for which trials would be prohibitively expensive. Recent efforts to control the Federal deficit will impact the amount of funding available for biomedical research. The 1% decrease in the National Institutes of Health budget for Fiscal Year 2011 was only the second reduction in the NIH budget since 1970. As research budgets continue to shrink in real dollar terms, the lower cost of observational methods will be increasingly attractive, especially for research questions in which there is no commercial enterprise willing to fund trials. For similar reasons, the larger sample sizes obtainable in observational research also facilitate clinically-relevant subgroup analyses that would not be adequately powered in a trial.

3. Finally, the use of observational data allows for comparisons that would be impractical or unethical in trials. Policy questions concerning the optimal benefits structure of government health insurance can be challenging to test using randomized trials for political reasons, particularly when many populations covered by government health insurance are considered to be vulnerable. Similarly, for those interventions recommended in

practice guidelines, pre-existing evidence of benefit can create an ethical prohibition against randomizing patients to a control group. Such ethical limitations, while clearly appropriate, nonetheless mean that it is difficult to use trials to measure the impact of non-adherence to clinical guidelines on patient outcome.

The three studies outlined in this dissertation leverage observational methods to conduct CVD-related health services research intended to inform efforts in the clinical and policy realms to improve outcomes while controlling costs. All three studies focus on questions related to the management of patients receiving percutaneous coronary intervention (PCI). Studies 1 and 2 focus on post-PCI care by characterizing patterns of care and outcomes associated with the use of coronary stress testing. Understanding the optimal use of cardiac testing procedures is an important clinical and policy problem, but a lack of evidence regarding the effects of testing on patient outcomes and treatment costs complicates these efforts. Study 3 focuses on a peri-procedural treatment decision – whether to deploy bare metal or drug-eluting stents. Each study is briefly described below.

Study 1: Stress testing patterns and predictors after elective PCI. Current treatment guidelines recommend that, for patients receiving elective PCI, stress testing should be reserved for the evaluation of recurrent symptoms rather than screening asymptomatic patients for recurrent ischemia (at least in the first two years following PCI). In this aim, we evaluated: (a) patterns in the use of stress testing after elective PCI; (b) predictors of stress test use; (c) correlations between facility-level variation in stress test use and both underling patient risk and subsequent patient outcomes.

Study 2: Stress testing imaging modalities after percutaneous coronary intervention. Another important area for which evidence regarding stress test choice is lacking is the

optimal imaging modality to employ when performing imaging-based exercise stress testing after PCI – echocardiography (cardiac ultrasound) or nuclear imaging. In this aim, we used longitudinal claims data from a population of Medicare beneficiaries receiving PCI for acute coronary syndromes to measure the downstream resource use and clinical outcomes associated with the choice of imaging modality.

Study 3: Local instrumental variables methods for evaluating drug-eluting and bare metal stents for PCI. Instrumental variables methods have gained some traction in the outcomes research literature, based on their ability to adjust for endogenous treatment selection (i.e., confounding) due to factors unobserved in the data available for analysis. However instrumental variables methods have a major limitation – they result in an estimate that reflects the *local average treatment effect* – the effect of a treatment on a hypothetical population that cannot generally be identified in practice. Methods developed by James Heckman and Edward Vytalcil and promoted in the outcomes research literature by Anirban Basu integrate an economic model of human behavior into the estimation approach in order to obtain results with greater policy relevance then the local average treatment effect available from traditional instrumental variable methods. Specifically, the use of "local instrumental variable" methods provide estimates of the average treatment effect, the effect of treatment on the treated, the effect of treatment on the untreated. In this study, I demonstrate an application of this methodology using the example of drug-eluting and bare metal coronary stents.

In concert, these three studies use high-quality, contemporary empirical approaches to answer questions of clinical and policy interest for patients living with CVD. Studies 1 and 2 inform the optimal use of stress testing. Non-invasive cardiovascular tests, including stress

testing, are commonly performed procedures and an important component of overall medical spending. Study 3 provides information about the comparative safety and efficacy of drug-eluting versus bare metal coronary stents. Perhaps more importantly, Study 3 also continues efforts to make the results of observational outcomes research more useful to providers, patients, and policymakers by promoting the use of newer methodologies that permit the estimation of intuitive estimands from instrumental variables models.

2. STRESS TESTING UTILIZATION AND MODALITY AFTER CORONARY REVASCULARIZATION

Understanding the optimal use of noninvasive cardiac testing procedures is an important clinical and policy problem, but a lack of evidence regarding the effects of testing on patient outcomes and treatment costs complicates these efforts. Papers 1 and 2 of this dissertation evaluate the use of stress testing after percutaneous coronary intervention. In this chapter, I provide additional background information on the use of coronary stress testing, its role in patients who have received elective PCI, and choices of stress testing modalities.

AN IMPORTANT POLICY PROBLEM

Non-invasive cardiac testing has undergone tremendous technological innovation in the past twenty years, as advancements across testing modalities have led to improvements in capabilities to evaluate cardiac physiology and anatomy while sparing patients the risks and expense of invasive tests such as cardiac catheterization. Enormous growth in utilization of non-invasive cardiac testing has resulted **(Figure 2.1)** (21, 22). The overall utilization and cost of medical imaging more than doubled between 2000 and 2005 with over $14 billion in Medicare Part B costs alone; one-third of medical imaging was used for cardiovascular applications (23). Despite this substantial increase in costs, the benefits of increased testing to patients are uncertain (22, 24). Marked geographic variation in rates of cardiac imaging use has been noted that cannot be explained by differences in underlying patient risk (25). Additionally, the increasing prevalence of cardiologists owning advanced imaging equipment

and strong associations between having an ownership stake in an imaging device and the rate at which cardiologists employ imaging have raised concerns that financial interests rather than medical necessity drive imaging decision making (22, 24, 26-28). The financial toll of the growth in testing use, in concert with evidence of geographic and provider variation in testing utilization, explains why non-invasive cardiac testing is an important focus for both clinicians and policymakers (29, 30).

Numerous efforts have been implemented to control unnecessary utilization of non-invasive cardiac testing. Payers have attempted to limit utilization through both changes in financial incentives and direct supervision of medical decision-making. Medicare reimbursement for cardiac imaging services has been reduced on several occasions – most recently in 2010, when CMS imposed a 36 percent reduction in reimbursement for the most common type of nuclear cardiac imaging test (single-photon emission computed tomography) and a 10 percent reduction for most forms of echocardiography (cardiac ultrasound) (31). In the private insurance market, there has been increased enthusiasm for applying managed care principles to imaging through pre-authorization review requirements administrated by radiology benefits management organizations (RBMs) (32, 33). The cardiology profession and allied groups have also responded to concerns about the growth in cardiac imaging use by developing voluntary practice guidelines from the American College of Cardiology and American Heart Association as well as Appropriate Use Criteria (AUC) from the American College of Cardiology Foundation (ACCF) (34).

Recent evidence suggests that this multifaceted approach is succeeding, with modest reductions in imaging utilization among Medicare beneficiaries **(Figure 2.2)** (35). However, each approach has weaknesses. While simple to implement, reimbursement changes are blunt

tools that do not distinguish between high- and low-value applications. There is evidence that cardiac imaging is overused for some applications, but also underused for some high-value indications (36). Thus, simple reimbursement reductions may have unintended consequences. The ACCF's AUC provide for clinical scenarios based on expert assessment on whether imaging use is appropriate (34). AUC have now been issued for all forms of cardiac imaging (37-39). However, AUC and similar guidelines committees frequently rely on expert opinion–only 1% of ACC/AHA imaging guidelines are based on "A"-quality evidence (40). It is common for experts to disagree – for example the routine use of echocardiography after cardiac resynchronization therapy implantation for device optimization was considered of uncertain benefit by AUC evaluators, with estimates of its appropriateness ranging widely from 4-9 across a fifteen judge panel (1-3 is considered "inappropriate", 4-6 "uncertain", and 7-9 "appropriate") (37). Thus, promulgation and widespread adoption of these guidelines is hampered by a paucity of rigorous empirical evidence regarding imaging's effects on outcomes, overall care processes, and cost (41, 42).

In response, the Agency for Healthcare Research and Quality (AHRQ) funded a contract with the Duke Clinical Research Institute (DRCI) through the Developing Evidence to Inform Decisions about Effectiveness program to develop new insights into the current patterns, predictors, and implications of stress testing after PCI (Project ID: 24-DKE-3; Work Assignment Number: HHSA290-2005-0032-I-TO4-WA3). This work has already led to insights about the use of coronary computed tomographic angiography after PCI (43) as well as the patterns and predictors of both stress test and angiography use after PCI (44). Studies 1 and 2 of this dissertation extend this work to additional questions pertaining to the use of stress testing in patients who have received PCI.

PATTERNS AND PREDICTORS OF STRESS TESTING AFTER ELECTIVE PCI (STUDY 1/CHAPTER 3)

Stress testing is a non-invasive test used to assess for significantly occluded coronary arteries. It is considered a “functional” test in that it measures the impact of occlusion on the myocardium (cardiac muscle), rather than an “anatomic” test (such as coronary angiography or cardiac computed tomographic angiography) that measures the narrowing of blood flow in the coronary arteries directly. The use of stress testing in patients who have previously received revascularization with PCI (commonly known as angioplasty) is an indication for which additional evidence is sorely needed to inform clinical practice. For patients who have received PCI, current major society guidelines and AUC consider assessment of recurrent chest pain to be an appropriate use of stress testing (referred to throughout this document as "symptom-driven testing"). However, the guidelines are also unanimous in recommending against the routine use of stress testing in asymptomatic patients to detect recurrent, asymptomatic cardiac ischemia after revascularization (referred to throughout this document as "surveillance testing") (38, 45, 46). This guidance was recently reflected by the American College of Physician’s “Choosing Wisely” campaign, in which the use of stress testing for surveillance testing was selected by both the American College of Cardiology and the American Society of Nuclear Cardiology as part of their lists of five common practices that ought to be questioned by patients and providers (47).

The consensus that surveillance testing is not an appropriate use of stress testing is primarily based on the evolving understanding of coronary artery disease pathology. Myocardial infarction and sudden cardiac death are generally caused by rupture of unstable coronary plaques. Unfortunately, most unstable plaques do not cause significant coronary

stenosis and consequently do not impact myocardial function until they rupture, thus they cannot be detected by stress testing (48, 49). The overwhelming number of lesions that are identified by surveillance stress testing would be those for which preemptive treatment with revascularization would be unlikely to prevent major cardiac events. While some of these lesions may eventually cause stable anginal symptoms, early intervention is unlikely to offer benefit. Furthermore, falsely positive results from stress testing subject patients to unnecessary anxiety and additional unnecessary procedures such as coronary angiography – procedures that come with their own attendant risks and costs.

While experts agree that stress testing should not be routinely performed after PCI, evidence from a large private insurer and the Medicare program demonstrated that stress testing is widely employed in the post-PCI population at rates that far exceed estimates of the incidence of recurrent chest pain in such a population (44, 50). These findings suggest that physicians are employing stress testing in post-PCI populations for indications other than those recommended by current AUC and clinical guidelines (*i.e.*, they are using stress testing for surveillance testing). Because the current guidelines provide only limited consideration of the likely heterogeneity in the value of stress testing, one possible explanation for the discordance between guidelines and practice is that physicians selectively use surveillance testing among patients who are expected to benefit more from it than the average patient. While small-scale registry and single center studies suggest that this is not the case (51, 52) and randomized trials of surveillance testing in populations enriched by design to include more patients with high risk characteristics have shown no benefit from such an approach (53-55), a large, nationally representative cohort study will provide more conclusive insights.

CHOICE OF STRESS TESTING MODALITY AFTER PCI (STUDY 2/CHAPTER 4)

While efforts continue to define more precisely *when* stress testing should be employed after PCI, understanding *which modality of* stress testing is optimal is also an important clinical and policy question. Two main parameters define how a stress test is conducted **(Figure 2.3)**:

1. Imaging modality: Stress testing can be conducted with electrocardiography (ECG) alone or with imaging; if imaging is used, by far the two most commonly employed techniques in the United States are nuclear imaging and echocardiography. No single approach is universally optimal. ECG-based stress testing is inexpensive, but has limited sensitivity and is often not interpretable in the setting of left bundle branch block or paced ventricular rhythms (46). While stress echocardiography and nuclear stress imaging perform similarly in ischemia detection, they differ in other key respects (56, 57). Fundamentally, echocardiography and nuclear imaging measure related, but subtlety different, phenomena. Nuclear imaging measures the perfusion of blood within the myocardium, and echocardiography identifies a "downstream" effect of low myocardial perfusions (motion defects). As a result, echocardiography has generally been shown to be less sensitive but more specific than nuclear imaging (58). Other than performance parameters, there are other considerations that may drive test choice. Stress echocardiography allows the concomitant identification of structural heart disease and avoids patient exposure to ionizing radiation; however, results are heavily dependent on the skill of both the technologist and interpreting physician as well as the patient's body habitus (59). Obtaining and interpreting results from nuclear stress testing is less subjective and modern nuclear techniques allow for the precise localization of myocardial ischemia, but nuclear stress testing is more expensive than other

modalities. In particular, the most common nuclear imaging technique (tomographic myocardial perfusion imaging) costs Medicare an average of $478, while stress echocardiography costs $210 (60). Recent evidence suggests that most stress testing conducted after revascularization is conducted with nuclear imaging; whether this approach represents optimal practice is unclear (50).

2. *Stress modality:* Stress testing can be conducted with stress induced by either exercise or administration of pharmacologic agents. Exercise stress testing provides information about cardiac function in the context of normal activity and avoids side effects from pharmacologic stress agents. In addition, exercise capacity is itself a useful prognostic indicator, and in elderly patients may be the strongest available predictor of mortality and cardiovascular events (61, 62). Stress induced by pharmacologic agents can only be performed with imaging (i.e., no ECG-only testing). Therefore, it is recommended that pharmacologic stress testing be reserved for patients who are unable or unwilling to exercise adequately (>85% of age-predicted maximal heart rate),(46) and guidelines for stress testing assume that providers will use exercise testing whenever possible (37, 38).

Preliminary data suggest the most promising comparison is between exercise echocardiography and exercise nuclear testing. Federspiel et al. conducted a retrospective cohort study using 2006-2008 data to identify contemporary patterns, predictors, and implications of stress testing modality in the year following PCI.(63) We made three binary comparisons of modality:

1) exercise testing with ECG alone versus exercise testing with imaging (either nuclear imaging or echocardiography);

2) pharmacologic stress testing with imaging versus exercise stress testing with imaging; and

3) exercise echocardiography versus exercise nuclear testing.

Using a combination of registry and administrative (claims) data from the Medicare program, we documented several findings:

1. Among patients receiving exercise stress tests, baseline clinical characteristics of patients receiving testing with and without imaging were similar. However, the proportion of tests performed without imaging was far higher in the first six months post PCI than later, suggesting that ECG-only exercise tests are employed for a different purpose than imaging-based exercise tests (i.e., that there is confounding by test indication) **(Figure 2.4)**[1]. Among patients receiving an imaging-based stress test, patients receiving pharmacologic stress testing were older and had higher rates of most comorbidities than patients receiving exercise stress testing, suggesting that this decision largely reflects differences in underlying risk (i.e., that there is confounding by patient characteristics). In contrast, patients receiving an exercise stress test with nuclear imaging appeared similar in most respects to patients receiving echocardiography, and there were no strong differences in test timing noted between echocardiography and nuclear-based tests. **Based on these results, I concentrated this Study on the decision to use exercise echocardiography or nuclear testing.**

2. Both with and without statistical adjustment for patient characteristics, there was pronounced geographic variation in whether patients undergoing an exercise stress test with imaging received nuclear or echocardiography **(Figure 2.5)**. Rates varied from 9.1% in the South Atlantic Census Division to 31.2% in the Pacific Division, a 3.4-fold difference.

---

[1] Figures 2.4 and 2.5 are reprinted from Federspiel JJ, Mudrick DW, Shah BR, Stearns SC, Masoudi FA, Cowper PA, et al. Patterns and predictors of stress testing modality after percutaneous coronary stenting: Data from the NCDR((R)). JACC Cardiovasc Imaging 2012;5(10):969-80, with permission from Elsevier

3. The choice of echocardiography versus nuclear imaging for exercise stress testing was associated with short-term differences in downstream procedure use. The cumulative incidence of repeat stress testing within 90 days of the initial stress test was higher in exercise echocardiography compared with exercise nuclear imaging tests, but the incidence of subsequent cardiac catheterization and revascularization were lower.

4. While the overall cumulative incidence of stress testing post-PCI declined by approximately 17 percent between 2006 and 2008, the proportion of stress tests performed with each modality varied little over the time period. These results suggest that broad-based reductions in stress test utilization post-PCI, rather than targeted changes in the use of specific modalities, have occurred in recent years.

Together, these results demonstrate broad variation in practice patterns in terms of whether patients receiving an exercise stress test with imaging receive echocardiography or nuclear imaging. This variation does not appear to be driven by differences in patient characteristics or test indications, but appear to be associated with differences in short-term procedure use after stress testing. The evidence from these preliminary findings and other previous studies on the competing advantages of echocardiography and nuclear imaging motivates the value of an analysis evaluating the relationship between variation in imaging modality choice and the outcomes and costs of stress testing.

FIGURES

Figure 2.1: Temporal trends in stress testing utilization compared with rates of invasive testing (catheterization), coronary revascularization and myocardial infarction: 1993-2001.

Source: Lucas 2006 (used with permission)

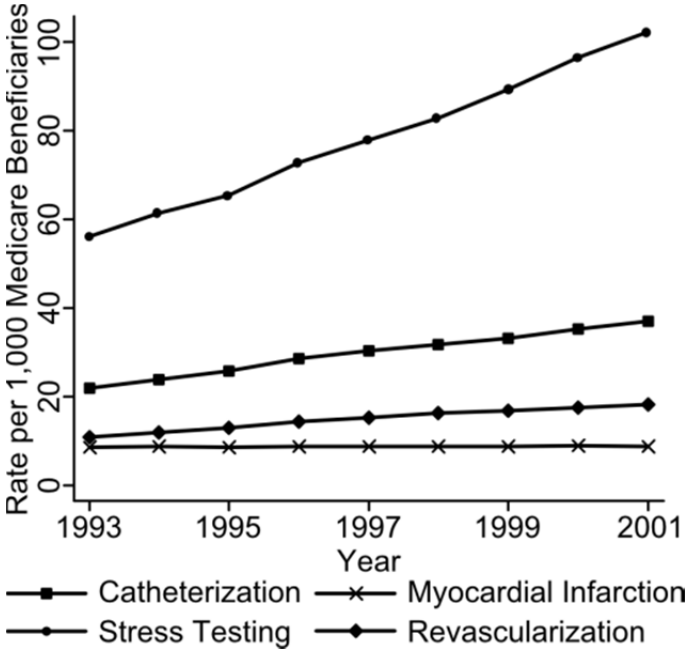


Figure 2.2: Temporal trend in Medicare Part B spending on nuclear imaging (MPS), echocardiography (Echo), exercise testing (ETT), and cardiac catheterization (Cath) services: 2000-2008.

Source: Shaw 2009 (used with permission)

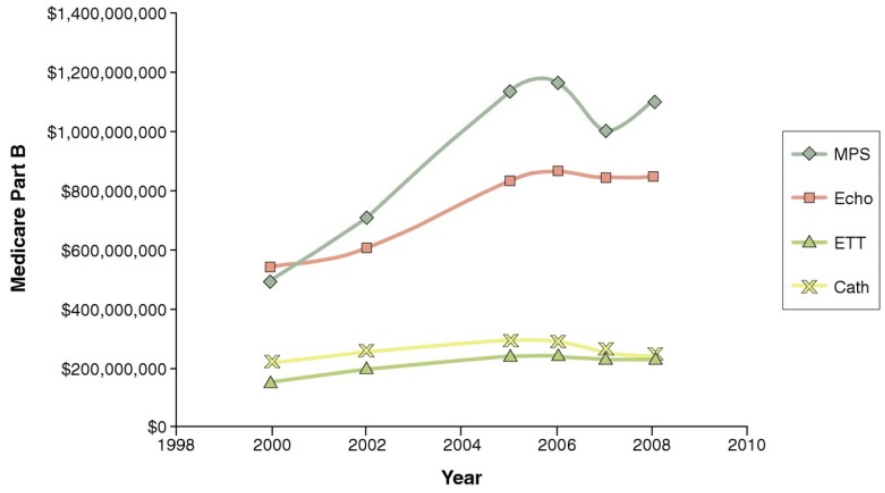


Figure 2.3: Forms of stress testing commonly performed in the United States

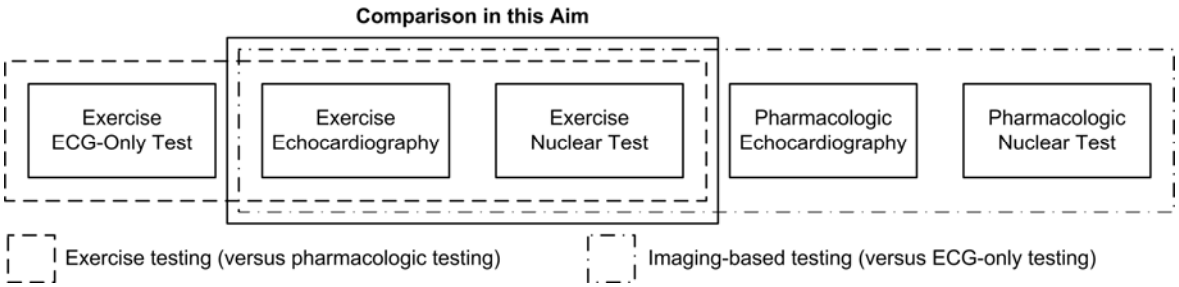


Figure 2.4: Trend in type of stress test employed as function of time since PCI.

Source: Federspiel 2012(Used with permission)

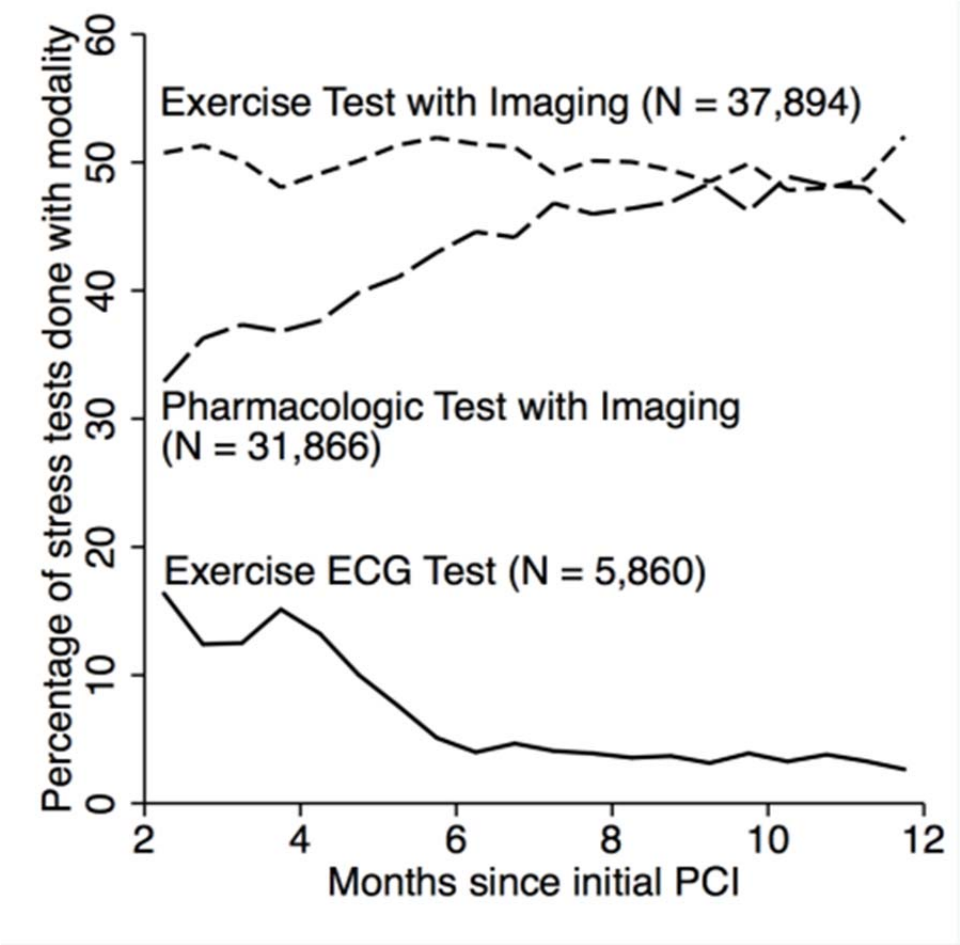
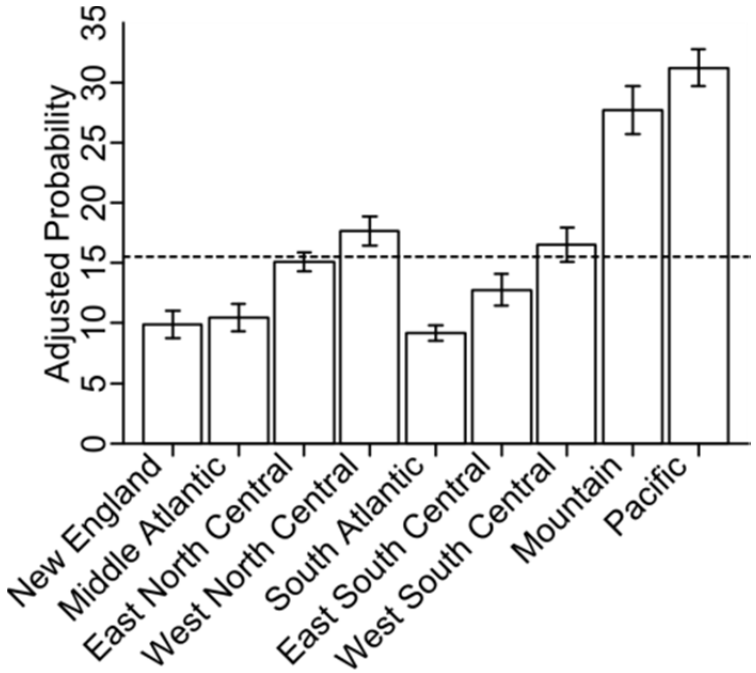


Figure 2.5: Probabilities of receiving an exercise stress testing with echocardiography rather than nuclear imaging, by US Census division.

Probabilities adjusted for patient, PCI procedure, and facility characteristics. Dotted line indicates national mean. *Source: Federspiel 2012*



3. STUDY 1: STRESS TESTING AFTER ELECTIVE REVASCULARIZATION

OVERVIEW

While current appropriate use criteria and guidelines consider stress testing not routinely indicated for surveillance in patients who received elective PCI for coronary artery disease, such testing may benefit higher-risk patients. Study objectives were to identify whether: (1) current practice patterns of stress testing after elective PCI reflect greater use in higher-risk patients, (2) facility variation in testing patterns reflects difference in underlying patient risk, and (3) facility variation in testing is associated with differences in clinical outcomes. This was a retrospective, observational cohort study using the national CathPCI Registry® linked to Medicare fee-for-service inpatient and outpatient claims for beneficiaries ≥65 years receiving elective PCI from 2005-2008. Outcome measures were cumulative incidence and timing of stress testing after elective PCI; facility-level variation in stress testing and clinical outcomes (death, readmission for myocardial infarction, repeat revascularization). Among 62,694 patients, stress testing incidence was 45.5% at 15 months post-PCI and 58.2% at 27 months. Among patients receiving stress testing, 51.8% received another within 27 months of the first. Patient-level factors associated with increased stress testing included an abnormal noninvasive study prior to the index PCI; negative predictors included no symptoms at time of index PCI, diabetes mellitus, incomplete revascularization, and ejection fraction <50%. In higher risk patients (history of silent ischemia, multi-vessel coronary disease, history of MI, diabetes mellitus, or receipt of incomplete revascularization) facility-level analysis

demonstrated no association between stress testing rates and death (p=0.61) or readmission for MI (p=0.76), but a positive association with repeat revascularization (p=0.008). Stress testing after elective PCI is common, performed more frequently in lower-rather than higher-risk patients, and occurs in a pattern indicating routine annual testing. Even in higher risk patients, associations between facility-level stress testing rates and increased use of repeat revascularization without differences in mortality or MI-related readmission suggest limited value from scheduled testing, providing an opportunity to improve management approaches after elective PCI.

INTRODUCTION

Current American College of Cardiology (ACC) Foundation appropriate use criteria (AUC) deem imaging-based stress testing within two years of percutaneous coronary intervention (PCI) as appropriate in patients who have recurrent symptoms consistent with further coronary obstruction ("symptom-driven testing"), but inappropriate in asymptomatic individuals ("surveillance testing") (37, 38). Routine stress testing was also deemed not beneficial in joint ACC/American Heart Association/Society for Cardiovascular Angiography and Interventions guidelines for PCI (64) and targeted as unnecessary in the ACC contribution to the "Choosing Wisely" initiative (47).

Despite this guidance, stress testing is performed after coronary revascularization in the United States at rates far exceeding the expected rate of recurrent symptoms of restenosis; suggesting widespread surveillance testing (50, 65). It is possible current testing patterns are driven by providers targeting higher-risk patients for surveillance testing, and that these patients do benefit from testing (*i.e.*, that the effect of routine stress testing is heterogeneous)

- limiting the value of broad recommendations such as the current AUC and clinical guidelines and motivating the value of targeted examinations of stress test usage.

To explore these issues, we linked detailed clinical data provided by the CathPCI Registry to longitudinal data from the Centers for Medicare & Medicaid Services (CMS) to determine among patients receiving elective PCI for coronary disease: 1) patterns of stress testing after elective PCI; 2) associations between clinical factors and stress testing use; and 3) for higher-risk patients, associations between facility-level utilization of stress testing after PCI and outcomes.

METHODS

Data Sources and Cohort

Cases of PCI with stent insertion were identified from the CathPCI Registry (66, 67), which is an initiative of the American College of Cardiology Foundation and The Society for Cardiovascular Angiography and Interventions. Details regarding dataset construction have been previously published (43, 63). Briefly, we included patients 65 years of age or older who were admitted and discharged between January 2005, and December 2008. Since the CathPCI Registry does not include direct identifiers, registry records were linked to fee-for-service Medicare inpatient claims using indirect identifiers (68, 69), which has been previously shown to produce a cohort representative of both the overall CathPCI Registry and Medicare populations aged 65 and older (70). For matched patients, we obtained CMS data from 2004-2008 that included outpatient and carrier (physician/supplier) claims and Medicare denominator files to enable identification of resource use and clinical events

subsequent to the index PCI procedure. Up to four years of follow-up data were available (i.e., from time of PCI through the end of 2008).

For this analysis, the cohort was restricted to individuals receiving elective PCI for a non-acute coronary syndrome indication, based on CathPCI Registry data. Additional exclusion criteria were applied to accurately measure subsequent stress testing and repeat revascularization and identify factors associated with stress testing. We excluded any stress testing done during a 60-day "blackout period" after each patient's index event, since diagnostic tests during this period may be routinely performed for cardiac rehabilitation, staging of procedures, or functional capacity assessments (50). We also excluded patients who ceased to be enrolled in fee-for-service Medicare, died, underwent repeat revascularization or angiography, or were readmitted for myocardial infarction (MI) during the blackout period.

Identification of Outcomes, Stress Test Use, and Covariates

Enrollment and mortality data were obtained from Medicare denominator files, while repeat coronary angiography and revascularization (PCI and coronary artery bypass grafting) events were identified using International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) procedure and Healthcare Common Procedure Coding System (HCPCS) codes from inpatient, carrier, and outpatient claims. Readmissions for MI were identified from ICD-9-CM primary diagnosis codes in inpatient claims.

Cardiac stress testing after PCI was identified by HCPCS codes on carrier or outpatient claims. Electrocardiogram (ECG) stress and nuclear imaging procedures performed within one day of each other were considered a single stress nuclear event (71). Stress ECG, and either echocardiography, cardiac magnetic resonance, and positron emission

tomography testing performed on the same day were combined into a single stress imaging event (see **Table 3.1** for codes).

Standard data definitions used in the CathPCI Registry were used for patient, procedural, and facility characteristics. One exception was revascularization completeness, which is not assessed in the CathPCI Registry (72). Therefore, in order to define revascularization completeness, registry data elements denoting maximum pre-PCI occlusion were consolidated into four coronary regions: left main, left anterior descending, circumflex, and right. Each region was considered to have a significant occlusion if maximum pre-PCI occlusion was reported to be ≥70% (≥50% for left main). We defined complete PCI as occurring when: 1) for each region in which significant occlusion was reported, ≥ 1 lesion (main vessel or major branch) was treated for pre-PCI occlusion reported ≥70% (≥50% for left main) and the post-PCI occlusion was reported as <50%; and 2) for all main vessel and major branch lesions treated in the region, post-PCI occlusion was reported as <50%. Because of the complexities inherent in assessing completeness of revascularization in bypass grafts, patients with a history of coronary artery bypass grafting were excluded. Because socioeconomic status was also not recorded in the CathPCI Registry, we obtained aggregate data at the patient ZIP code level from the 2009 Pop-Facts database (Nielsen Claritas, Ithaca, New York).

Statistical Analyses

Patterns of Stress Test Use

Receipt of stress testing after the 60-day blackout period was assessed using cumulative incidence functions treating death, readmission for MI, repeat revascularization, and repeat

angiography as competing risks to identify the use of stress testing in patients without a major cardiac event after PCI; termination of fee-for-service Medicare coverage and end of claims availability were treated as censoring events. Testing patterns are presented both as cumulative incidence curves portraying the percentage of patients (accounting for censoring) who received a stress test up to a particular time point, as well as kernel-smoothed estimates of the hazard showing the rate at which stress testing was employed (the "intensity of testing") at a specific time point (73). In addition to evaluating the use of first stress tests after PCI, second and third stress tests were also assessed to characterize longer term patterns of repeated testing.

Clinical Correlates of Stress Test Use

Characteristics of patients undergoing elective PCI were described using medians, 25th and 75th percentiles for continuous variables and percentages for categorical variables. Associations between individual characteristics and stress testing (not preceded by death, MI-related readmission, or repeat revascularization or angiography) were calculated with both unadjusted and adjusted cause-specific Cox proportional hazards models (74). In addition to adjusting for patient and procedural characteristics, we also constructed a model which included fixed effects for each PCI facility (using stratification rather than indicator variables to avoid the incidental parameter problem); this approach allowed for adjustment for differences in stress test use attributable to unobserved facility-level differences in patient risk or practice patterns. For selected covariates, we compared the unadjusted cumulative incidence of stress testing at 27 months after PCI (approximately two years after PCI) using Gray's test (75). To evaluate whether clinical characteristics were associated with differences

in overall stress testing use and timing, we constructed extended proportional hazards models that included interaction terms between covariates and a binary indicator for time being in the seventh, thirteenth, and fourteenth months after PCI. These time points were selected based on inspection of testing rates in the data as appearing to correspond to follow-up visits at 6 and 12 months post-PCI. We hypothesized that a different effect of a characteristic on stress testing use in these months versus the remainder of the follow-up period would suggest a different effect on surveillance stress testing use than symptom-driven testing, since surveillance testing is more likely to occur periodically at regular intervals and associated with interval office visits than symptom-driven testing. Wald tests were used to assess the significance of the interaction terms.

Facility-level variation in test use in higher-risk individuals and patient outcomes

The cohort was limited to higher-risk patients who may have greater benefit from surveillance stress testing: history of silent ischemia (asymptomatic at PCI), multi-vessel coronary disease, history of MI (prior to their elective index PCI), diabetes mellitus, and receipt of incomplete revascularization. Because only limited data were available about patient's health status after discharge from PCI, we conducted an indirect analysis leveraging variation in facility rates of stress testing, (calculated using cause-specific stress testing incidence 15 months after PCI, based on the facility in which the patient received PCI). Facilities were categorized into quartiles based on stress testing use; only facilities with ≥25 patients were included to reduce measurement error.

Cause-specific Cox proportional hazards models were used to model the association between calculated facility-level quartile of stress testing and patient outcomes (all-cause

mortality, MI-related readmission, repeat coronary revascularization). All models considered end of fee-for-service claims availability to be a censoring event; for the MI endpoint, death was another censoring event; for repeat revascularization, both death and MI were considered to be censoring events. In addition to the facility-level quartile of stress testing, models for all three outcomes adjusted for patient, procedural, and facility characteristics specified *a priori* as possibly associated with differences in outcome. A marginal model was used to account for the likely correlation in error terms within-facility, with standard errors calculated using a robust sandwich variance estimator clustered at the facility level (74). We assessed the robustness of results with respect to model choice in sensitivity analyses in which we: 1) obtained standard errors via bootstrapping, sampling at the facility level, and 2) incorporated facility-level random effects to estimate a shared frailty (hierarchical or cluster-specific) model instead of a marginal (population-averaged) model (76).

This indirect approach assumed that variations in stress testing use across facilities would reflect variations in practice patterns- that high utilization rates suggest more frequent use of surveillance stress testing, and that low utilization rates would suggest stress testing reserved for recurrent symptoms. To assess the validity of this assumption, we calculated kernel-smoothed hazard rates, based on facility quartile of stress test use, to determine if facilities with a higher rate of stress test use had a more strongly periodic pattern of stress testing usage (spikes in stress testing utilization at time points suggesting surveillance testing at scheduled intervals). We also assessed whether underlying risk was correlated with variation in facility-level rates of stress testing usage. Given the large sample size, statistically significant differences in individual characteristics across quartiles of stress test use may not be clinically meaningful, and may be counter-balanced by differences in other

characteristics. To assess the net effect of observable differences in characteristics on expected outcomes, we constructed cause-specific Cox models, which included all covariates except the facility quartile of stress testing. Using this model, we calculated predicted relative hazards for each patient, and averaged them across facility quartiles of stress testing, with standard errors provided by the delta method. Higher rates of stress testing may alternatively indicate that patients receive stress testing prior to repeat angiography procedures, with lower rates of stress testing reflecting facilities in which patients with recurrence of symptoms are referred directly to repeat angiography. Consequently, as sensitivity analysis we also estimated a proportional hazards model including facility-level quartiles of combined repeat angiography or stress testing.

Statistical analyses were conducted using SAS version 9.2 (SAS Institute, Cary, North Carolina), Stata/SE version 12.1 (Statacorp LP, College Station, Texas), and R version 2.11.1 (R Foundation, Vienna, Austria), using an alpha level of 0.05. The Duke University Medical Center and University of North Carolina Institutional Review Boards granted a waiver of informed consent and authorization for this study; all analyses were performed at the Duke Clinical Research Institute (Durham, NC).

RESULTS

Patient Cohort

Overall, 62,964 patients treated at 875 centers underwent elective PCI between January 1, 2005, and October 31, 2008 and met criteria for the 60-day blackout period **(Figure 3.1)**. The median duration of follow-up was 625 days [25^{th}–75^{th} percentile: 323–960 days]. The median age was 74 years, 56.8% were men, and 87.4% were white.

Patterns of Stress Test Use

By four years after PCI, 66.4% of patients received at least one stress test, while 19.7% experienced a major coronary event prior to stress testing (a "competing event") **(Figure 3.2A)**. Almost all (94.1%) of the initial post-PCI stress tests were performed with imaging, primarily nuclear imaging (84.4%). Stress testing in the first two years after PCI was common: the cumulative incidence of stress testing was 45.5% and 58.2% by 15 and 27 months, respectively. Stress testing rates showed peaks in the 7th and 13-14th months after stress testing, suggesting frequent use of surveillance testing at these time points **(Figure 3.2B)**. Repeated stress testing was also common; by 27 months after an initial stress test, 51.8% of patients received a second **(Figure 3.2C)**; similarly, by 27 months after a second stress test, 58.7% received a third **(Figure 3.2E)**. Rates of second stress testing peaked in the 13-14th months after an initial stress test **(Figure 3.2D)**, and third stress testing rates demonstrated a similar pattern **(Figure 3.2F).** As with initial tests, the vast majority (94.1%) of repeat stress tests were performed with imaging. When CABG recipients were included in secondary analysis, results were very similar (not shown).

Predictors of Stress Test Use

The unadjusted cumulative incidence of stress testing at 27 months post-PCI was lower for patients who were asymptomatic at time of PCI, received incomplete revascularization, had diabetes, multi-vessel coronary disease, a history of MI, were ≥ 75 years, or had an ejection fraction <50% at time of PCI; it was higher for patients who had a positive pre-PCI ischemia test result **(Figure 3.3)**. Cox modeling, with addition of facility

fixed effects, demonstrated that after adjustment, factors positively associated with receiving a stress test at any time within 4 years of PCI **(Table 3.2)** included neighborhood-level rates of college education, statin therapy use at time of PCI, family history of coronary artery disease, positive ischemia test result prior to the index PCI procedure, and a history of multiple PCI procedures as well as receiving a drug-eluting stent or multiple lesions treated during the index PCI procedure, and receiving PCI at a private rather than university hospital. Independent negative predictors included age ≥ 75 years, most comorbid conditions (cerebrovascular disease, chronic lung disease, heart failure, and peripheral vascular disease), being a current smoker, having a history of prior MI (not at the time of the encounter or during the study period), being asymptomatic at time of PCI, incomplete coronary revascularization, and having a major PCI-related complication. While there was a modest difference between stress testing rates between white and nonwhite patients (HR 0.96, p = 0.032) prior to inclusion of hospital fixed effects, this effect was eliminated with addition of fixed effects (HR 0.99).

Next, we introduced interaction terms between the covariate and time (7^{th} or $13\text{-}14^{th}$ months post-PCI based on the pattern of stress testing in Figure 3.1B) **(Figure 3.4)**. Statistically significant interactions were observed for asymptomatic status at time of PCI ($p=0.01$), having a positive pre-PCI ischemia test ($p<0.001$), and ejection fraction <50% ($p=0.03$). Asymptomatic patients at the time of index PCI had a similar rate of stress testing as symptomatic patients during months 7 and 13-14, but a lower rate of testing in all other months. Compared with patients with no, negative, or equivocal pre-PCI ischemia testing, having a positive pre-PCI test was even more strongly associated with increased use of stress testing in months 7 and 13-14 than it was the remainder of the study period. Conversely,

while patients with ejection fraction <50% had lower rates of stress testing overall, this association was more strongly negative in months 7 and 13-14 than the remainder of the period.

Facility-level variation in test use in higher-risk individuals and patient outcomes

The higher-risk cohort included 43,512 patients. There was significant variation in the 15-month rate of stress testing usage after PCI, based on the facility in which the patient received PCI. Relative to the lowest test use quartile, the hazard ratio for stress testing within 15 months of PCI increased to 1.59 (CI, 1.50–1.69), 2.09 (1.97–2.21), and 3.12 (2.91–3.35) in quartiles 2-4, respectively. The pattern of stress testing also differed across quartiles **(Figure 3.5)**. Little periodicity was observed in the first quartile, while the fourth had a strongly periodic pattern, peaking in the 7th and 13-14th months post-PCI. This trend suggests that surveillance testing was commonly performed in the fourth quartile, while patients treated in facilities in the first quartile were unlikely to receive such surveillance testing. Statistically significant, albeit modest differences in many patient characteristics were observed across quartiles of stress testing **(Table 3.3)**. Based on characteristics in the dataset, patients who received PCI in facilities in the highest quartile had less cerebrovascular disease, heart failure, pre-PCI ischemia testing and more positive pre-PCI ischemia testing, compared to lower quartiles. These patients were, on average, at lower risk of death, MI, and repeat revascularization than those who were treated in the lowest quartile. In aggregate, on average patients receiving PCI at facilities in the highest quartile of stress testing rates after PCI were predicted to have 10% lower rates of death and MI, and approximately 4% lower rates of repeat revascularization, than those treated in lowest quartile facilities **(Table 3.4).**

Covariate-adjusted, cause-specific Cox models, coefficients for facility quartiles 2-4 (relative to quartile 1) were not jointly significant for death (p=0.61) or MI (p=0.76), nor were any pairwise comparisons to quartile 1 **(Figure 3.6)**. In contrast, for repeat revascularization, terms for quartiles 2-4 were jointly significant (p=0.008), with increasing use of repeat revascularization associated with the increasing quartiles of stress testing, and greatest increase for patients receiving index PCI in fourth (highest) quartile facilities. Similar results were observed: 1. with quartiles calculated based on the combined rate of angiography and stress testing; 2. with bootstrapped standard errors; and 3. when using a shared frailty model rather than a marginal model **(Table 3.5).**

DISCUSSION

In this national cohort of fee-for-service Medicare beneficiaries receiving elective PCI, subsequent stress testing was commonly performed for surveillance, rather than recurrent symptoms. Patient factors that would predict a high likelihood of restenosis or recurrent symptoms were paradoxically associated with lower rates of post-PCI stress testing. Associations between facility-level stress testing rates in the 15 months following elective PCI and clinical outcomes up to 4 years post-PCI demonstrated a positive relationship between stress testing rates and repeat revascularization in higher risk patients, but no association with respect to mortality or MI-related readmission.

Consistent with previous reports in privately-insured (50) and Medicare beneficiaries (65), we find high rates of stress testing in the years following elective PCI. Testing rates exceed the estimated incidence of recurrent angina in the year after PCI (18-20%) (77). Similar to previous analyses of non-Medicare beneficiaries (50), the timing of stress testing

suggests that initial surveillance tests are commonly performed at 6 and 12 months post-PCI. Our analysis extends previous results not only by its focus on elective PCI but also by evaluating patterns of repeated stress testing after PCI both overall, and in higher risk individuals. We find that over half of patients receiving an initial post-PCI stress test receive additional stress tests within two years, with timing indicative of annual surveillance testing. These results are consistent with other recent work in Medicare beneficiaries, which demonstrated a similar incidence and pattern of repeated stress testing (78).

Surprisingly, patient factors predicted to increase the risk of recurrent ischemia and presumably the value of routine testing (e.g., incomplete revascularization, diabetes mellitus, prior MI, increasing age, low ejection fraction) were associated with reduced use of stress testing after elective PCI. In addition, while often statistically significant, the magnitude of observed effects was modest – none altered the overall rate of stress testing by even 20%. This finding suggests that underlying patient risk may not be a major factor driving decisions regarding stress testing in patients who have received elective PCI.

The limited impact of patient risk factors on use of stress testing was striking when considering the wide variation in testing patterns observed in facilities where patients received PCI (a proxy measure for local practice patterns). Furthermore, some of the strongest predictors of increased stress testing were receipt of PCI in a private (versus university) facility and the neighborhood-level rate of college degree attainment. The reasons for these findings are unclear, but may reflect differences in perceived or actual malpractice risk (79), financial incentives from cardiologist ownership of imaging equipment or health system ownership of the cardiology practice (28), or local differences in patient preferences. Also of note was the lack of difference in testing patterns based on gender. There was a very

modest difference observed based on race, suggesting less post-PCI stress testing for white patients. However, observed race-based differences in care were eliminated with application of facility-level fixed effects, suggesting that differences in facility-level practice patterns explains the lower rate of stress test use in whites.

The observed variation in stress testing rates and periodicity even in a higher risk patient group based on the facility where the patient received PCI, suggests that higher-use facilities performed more surveillance testing, while low-use facilities reserved testing for patients with recurrent symptoms. Rather than increased stress testing rates reflecting increased patient risk, we observed greater use of stress testing in facilities where patients were, on average, at lower risk for death, MI, and repeat revascularization. Others have shown similarly broad facility-level variation in stress testing patterns and a lack of correlation between patient risk and test use (51, 52). Facility-level variation in stress testing was associated with increased rates of coronary revascularization, but not death or MI-related readmission. These results complement those from a recent single-center study of CABG and PCI recipients referred for screening stress echocardiography (53), small randomized trials in high-risk populations (54, 55), and our previous work in an "all-comers" Medicare PCI population.

Limitations

This large cohort study has many strengths, but we acknowledge several limitations. First, the cohort includes fee-for-service Medicare beneficiaries who are treated at facilities participating in the CathPCI Registry. While the merged dataset is representative of older Medicare beneficiaries and CathPCI Registry participants,(70) the generalizability of results

to younger patients is unknown. Second, data after PCI discharge are limited to Medicare claims. Consequently, we cannot know whether individual stress tests were consistent with AUC or practice guidelines - forcing the indirect analytic approaches employed here. Third, the validity of the facility-level variation analysis relies on the untestable assumption that after adjustment for observed factors, unobserved factors do not confound the relationship between facility-level stress test use and outcomes. Finally, patients may benefit from surveillance stress testing through outcomes not measured in this analysis (e.g., quality of life, angina, or reassurance).

Conclusion

In summary, our findings suggest that stress testing is commonly used for surveillance after elective PCI, a strategy that is not recommended by current AUC and guidelines. Clinical factors predicting risk of recurrent ischemia, and possibly increased testing value seem to have modest influence on the overall testing rate. Even in higher risk patients, greater use of stress testing was not associated with either death or MI readmission, but was associated with an increased rate of revascularization. This finding is consistent with emerging evidence and consensus-based guidelines that surveillance testing in asymptomatic patients does not improve outcomes, even in the higher-risk patient cohort we examined, and may expose patients to unnecessary procedures. The implementation and assessment of a surveillance mechanism using detailed clinical data and focusing on the appropriateness of testing after PCI has the potential to improve the value of management approaches in this high risk population.

TABLES

Table 3.1: Healthcare Common Procedure Coding System and International Classification of Diseases, Ninth Revision, Clinical Modification codes used to define stress testing and outcomes.

	HCPCS or ICD-9-CM Code	Type of Claim
Stress Testing		
Electrocardiogram stress test*	93015-93018	Carrier/Outpatient
Nuclear imaging	78460-78461, 78464-78465, 78472-78473, 78481, 78483	Carrier/Outpatient
Echocardiography	93350	Carrier/Outpatient
Positron Emission Tomography	78491-78492	Carrier/Outpatient
Magnetic Resonance Imaging	75552-75556 (2005-2007) 75559-75560, 75563-75564 (2008)	Carrier/Outpatient
Coronary Angiography	93508, 93539, 93540, 93545	Carrier, Outpatient
Percutaneous Coronary Intervention	92980-92982, 92984, 92995, 92996, G0290, G0291, 36.01, 36.02, 36.05, 36.06, 36.07, 00.66	Carrier, Outpatient, Inpatient
Coronary Artery Bypass Grafting	33510-33514, 33516-33519, 33521-33523, 33533-33536, 36.1x, 36.2, S2205-S2209	Carrier, Outpatient, Inpatient
Acute Myocardial Infarction	410.x1 (principal diagnosis)	Inpatient

*Electrocardiogram stress and nuclear imaging procedures performed within one day of each other were considered a single stress nuclear event. Electrocardiogram stress and all other imaging performed on the same day were considered a single stress imaging event.

Table 3.2: Selected characteristics at time of elective PCI, and associations with stress testing rate. (N=62,694)

	% or median [Q1–Q3]	Unadjusted Hazard Ratio (CI)	Adjusted* Hazard Ratio (CI)	Adjusted* Hazard Ratio, with Facility Effects (CI)
Demographics				
Age (years)				
Median [Q1 - Q3]	74 [69-79]	-	-	
75 years or older	46.2	0.81 (0.79, 0.83)	0.82 (0.80, 0.84)	0.80 (0.78, 0.82)
Male gender	56.8	1.05 (1.02, 1.07)	1.01 (0.99, 1.03)	1.01 (0.98, 1.03)
White Race	87.4	0.99 (0.96, 1.03)	0.96 (0.93, 1.00)	0.99 (0.95, 1.03)
ZIP Code Characteristics				
Mean household income ($10,000s)	6.0 [5.1-7.8]	1.04 (1.04, 1.05)	1.02 (1.01, 1.02)	0.99 (0.98, 1.01)
% of households below poverty line (HR per 100% difference)	6.9 [3.9-10.9]	0.27 (0.22, 0.33)	0.69 (0.49, 0.99)**	0.74 (0.49, 1.12)**
% of households renting house (HR per 100% difference)	23.2 [16.4-31.4]	0.90 (0.82, 0.98)**	1.01 (0.89, 1.14)**	0.89 (0.77, 1.03)**
% of adults with less than HS diploma (HR per 100% difference)	16.8 [11.1-23.9]	0.41 (0.36, 0.46)**	1.13 (0.85, 1.49)**	1.05 (0.75, 1.46)**
% of adults with a college degree (HR per 100% difference)	25.7 [18.6-38.0]	1.98 (1.83, 2.13)**	1.49 (1.21, 1.83)**	1.54 (1.21, 1.96)**
Clinical History				
Body mass index (HR per 10 units)	28.1 [25.0-32.0]	1.02 (1.00, 1.04)***	0.99 (0.97, 1.01)***	0.98 (0.96, 1.00)***
Comorbidities				
Cerebrovascular Disease	13.7	0.84 (0.81, 0.87)	0.89 (0.86, 0.92)	0.89 (0.86, 0.93)
Chronic Lung Disease	17.0	0.86 (0.84, 0.89)	0.94 (0.91, 0.98)	0.95 (0.92, 0.98)

Diabetes Mellitus	31.9	0.95 (0.93, 0.98)	0.97 (0.94, 0.99)	0.97 (0.94, 1.00)
Dyslipidemia, on statin	75.7	1.10 (1.07, 1.13)	1.07 (1.04, 1.10)	1.08 (1.05, 1.11)
Heart Failure	10.7	0.72 (0.69, 0.75)	0.84 (0.81, 0.88)	0.86 (0.82, 0.90)
Hypertension	82.4	0.98 (0.95, 1.01)	1.00 (0.97, 1.03)	0.99 (0.96, 1.02)
Peripheral vascular disease	14.1	0.86 (0.83, 0.89)	0.94 (0.90, 0.97)	0.93 (0.90, 0.96)
Current smoker	10.6	0.84 (0.81, 0.88)	0.85 (0.82, 0.89)	0.83 (0.80, 0.87)
Family history of CAD < 55 years	20.8	1.08 (1.05, 1.11)	1.07 (1.04, 1.10)	1.07 (1.03, 1.10)
Pre-PCI Ischemia Testing				
None	18.4	0.98 (0.93, 1.03)	0.97 (0.92, 1.02)	0.96 (0.91, 1.01)
Negative	9.9	1 (reference)	1 (reference)	1 (reference)
Equivocal	2.4	0.99 (0.91, 1.08)	0.99 (0.90, 1.08)	1.08 (1.03, 1.13)
Positive	69.2	1.19 (1.14, 1.24)	1.13 (1.08, 1.18)	1.00 (0.91, 1.10)
Previous PCI (prior to index PCI)	26.8	1.05 (1.02, 1.08)	1.30 (1.22, 1.39)	1.26 (1.18, 1.35)
Previous PCI > 1 year ago or unknown timing	23.8	1.02 (0.99, 1.05)	0.79 (0.74, 0.85)	0.79 (0.74, 0.85)
Previous MI (> 7 days)	18.6	0.89 (0.87, 0.92)	0.92 (0.89, 0.95)	0.93 (0.90, 0.96)
Left Ventricular Ejection Fraction < 50%	15.7	0.79 (0.76, 0.81)	0.88 (0.84, 0.91)	0.88 (0.84, 0.91)
Admission Symptoms				
Asymptomatic	36.3	0.95 (0.92, 0.97)	0.96 (0.94, 0.98)	0.93 (0.90, 0.95)
Atypical Chest Pain or Stable Angina	63.7	1 (reference)	1 (reference)	1 (reference)
Active heart failure at time of PCI	8.1	0.75 (0.72, 0.79)	0.96 (0.91, 1.01)	0.91 (0.86, 0.97)
Glomerular filtration rate (eGFR; mL/min/1.73m^2)				
eGFR < 30 or on dialysis	3.2	0.81 (0.75, 0.87)	0.94 (0.88, 1.02)	0.97 (0.90, 1.04)
30 ≤ eGFR < 60	33.5	0.92 (0.90, 0.94)	0.98 (0.95, 1.00)	0.97 (0.95, 1.00)
eGFR ≥ 60	63.3	1 (reference)	1 (reference)	1 (reference)

PCI Procedural Characteristics				
Incomplete revascularization	25.3	0.91 (0.89, 0.94)	0.96 (0.93, 0.99)	0.95 (0.92, 0.98)
Any drug-eluting stent used	78.1	1.15 (1.11, 1.18)	1.06 (1.03, 1.10)	1.08 (1.04, 1.11)
Number of lesions treated	1 [1-2]	1.02 (1.00, 1.04)	1.04 (1.02, 1.06)	1.03 (1.01, 1.06)
Multivessel Disease	40.0	0.94 (0.92, 0.96)	0.97 (0.94, 1.00)	0.92 (0.86, 0.98)
Any Procedural Complication	3.9	0.89 (0.84, 0.95)	0.92 (0.86, 0.98)	0.95 (0.92, 0.98)
Facility Characteristics				
Facility Type				
Private, teaching	39.2	1.16 (1.11, 1.21)	1.19 (1.14, 1.25)	N/A**
Private, non-teaching	48.9	1.24 (1.19, 1.30)	1.31 (1.25, 1.37)	N/A**
University	10.1	1 (reference)	1 (reference)	N/A**
Government	1.8	0.94 (0.85, 1.04)	0.91 (0.82, 1.01)	N/A**
Annual PCI Volume (HR per 100 procedures)	853 [550-1482]	1.00 (1.00, 1.00)	1.01 (1.00, 1.01)	N/A**
Bed Size (HR per 100 beds)	421 [299-576]	1.00 (0.99, 1.00)	1.00 (0.99, 1.01)	N/A**

* In addition to listed variables, also adjusted for Census Region and year of PCI.

** N/A because fixed effects removes time-invariant facility characteristics

Table 3.3: Patient characteristics (median [Q1,Q3] or percentage), by facility quartile of stress test use at 15 months after percutaneous coronary intervention

	Overall 472 Facilities (N = 43,512)	Quartile 1 (Lowest rate) 118 Facilities (N = 9,135)	Quartile 2 118 Facilities (N = 10,893)	Quartile 3 118 Facilities (N = 11,426)	Quartile 4 (Highest rate) 118 Facilities (N = 12,058)	p-value
Demographics						
Age (years)						
Median [Q1 - Q3]	74 [69, 79]	73 [69, 79]	74 [69, 79]	74 [69, 79]	74 [69, 79]	<0.001
75 years or older	46.3	44.5	46.9	46	47.6	<0.001
Female gender	41.4	42.8	43	40.8	39.6	<0.001
White Race	87.1	89.0	86.7	85.9	87.3	<0.001
ZIP Code Characteristics						
Mean household income ($10,000s)	6.0 [5.1, 7.8]	5.6 [4.9, 6.6]	5.6 [4.9, 7.2]	6.6 [5.3, 8.5]	6.4 [5.3, 8.5]	<0.001
% of households below poverty line	6.8 [3.9, 10.8]	7.8 [5.1, 11.0]	7.7 [4.3, 12.1]	5.9 [3.5, 10.2]	5.9 [3.4, 9.7]	<0.001
% of households renting house	23.2 [16.4, 31.4]	22.9 [16.5, 30.7]	23.8 [16.9, 31.0]	22.8 [15.9, 31.9]	23.3 [16.3, 32.4]	0.002
% of adults with less than a high school education	16.7 [11.1, 23.7]	18.1 [13.1, 24.6]	18.3 [12.4, 25.5]	15.6 [10.2, 22.7]	15.4 [9.8, 21.9]	<0.001
% of adults with a college degree	25.7 [18.6, 37.9]	22.9 [17.6, 32.2]	22.9 [17.1, 34.1]	28.8 [20.1, 41.1]	28.1 [20.4, 41.9]	<0.001
Clinical History						
Body mass index [kg/m2]	28.3 [25.2, 32.3]	28.5 [25.4, 32.7]	28.4 [25.1, 32.4]	28.2 [25.1, 32.1]	28.2 [25.2, 32.1]	<0.001
Comorbidities						
Cerebrovascular Disease	14.7	14.8	15.6	15.1	13.3	<0.001
Chronic Lung Disease	17.1	18	18.9	16.6	15.5	<0.001
Diabetes Mellitus	41.1	42.5	41.8	40.3	40.3	<0.001
Dyslipidemia, on statin	76.8	75.9	76.8	77.5	76.8	0.05

Heart Failure	11.8	12	13.2	12.1	10.1	<0.001
Hypertension	83.4	83.5	83.8	83.4	82.9	0.32
Peripheral vascular disease	15.2	14.3	16.9	15	14.5	<0.001
Current smoker	10.8	11.3	11.1	10.4	10.4	0.10
Family history of CAD < 55 years	20.6	22.4	20.3	20.5	19.6	<0.001
Pre-PCI Ischemia Testing						<0.001
None	17.7	19.4	20.4	16.4	15	
Negative	2.3	2.3	2.7	2.5	1.5	
Equivocal	9.8	11.9	11.3	9	7.6	
Positive	70.3	66.4	65.6	72	75.8	
Previous PCI (before index procedure)	28.9	26.9	28.6	28.7	30.9	<0.001
Previous PCI > 1 year before index procedure, or timing unknown	25.8	24.2	25.4	25.6	27.5	<0.001
Previous MI (> 7 days)	24.1	23.5	26.4	24.1	22.4	<0.001
Left Ventricular Ejection Fraction						
Median [Q1 - Q3]	60.0 [50.0, 63.0]	60.0 [50.0, 65.0]	60.0 [50.0, 61.0]	60.0 [50.0, 63.0]	60.0 [50.0, 60.0]	<0.001
< 50%	20.4	20.2	21.8	20.1	19.7	0.004
Asymptomatic at PCI	46.6	40.3	45.7	47.7	51.1	<0.001
Active heart failure at time of PCI	8.5	9.2	8.7	8.3	7.8	0.001
Glomerular filtration rate (mL/min/1.73m^2)						0.003
eGFR < 30 or receiving dialysis	3.6	4.1	3.7	3.6	3.2	
30 ≤ eGFR < 60	34.1	35.1	34.1	33.7	33.5	
eGFR ≥ 60	58.4	57.3	58.2	59.5	58.5	

PCI Procedural

Characteristics						
Incomplete revascularization	32.6	33.2	31.2	33.2	32.8	0.004
Any drug-eluting stent used	78.4	79.9	76.9	79.5	77.6	<0.001
Multivessel Disease	51.7	54.2	49.3	52.2	51.6	<0.001
Any Procedural Complication	4.1	3.8	3.9	4.6	3.9	0.006
Facility Characteristics						
Facility Type						
Private, teaching	40.0	42.6	43.5	37.1	37.6	<0.001
Private, non-teaching	48.0	36.9	45.8	47.2	59.2	<0.001
University	10.2	14.4	10.8	14.6	2.4	<0.001
Government	1.8	6.0	0.0	1.1	0.8	<0.001
Annual PCI Volume	921 [601, 1597]	877 [603, 1334]	1001 [632, 1668]	946 [593, 1699]	878 [567, 1572]	<0.001
Bed Size	431 [300, 589]	445 [343, 568]	385 [300, 560]	500 [323, 650]	399 [279, 551]	<0.001

Table 3.4: Comparison of expected event rates, across quartiles of stress test utilization

	Hazard Ratio (CI)	P
Death		<0.001
Quartile 1 (Lowest)	1.00 (Reference)	
Quartile 2	1.04 (1.01, 1.06)	
Quartile 3	0.94 (0.92, 0.97)	
Quartile 4 (Highest)	0.92 (0.89, 0.95)	
MI-related Readmission		<0.001
Quartile 1 (Lowest)	1.00 (Reference)	
Quartile 2	1.02 (0.99, 1.06)	
Quartile 3	0.94 (0.89, 0.98)	
Quartile 4 (Highest)	0.90 (0.85, 0.95)	
Repeat Revascularization		<0.001
Quartile 1 (Lowest)	1.00 (Reference)	
Quartile 2	0.98 (0.96, 1.00)	
Quartile 3	0.98 (0.95, 1.01)	
Quartile 4 (Highest)	0.96 (0.93, 1.00)	

Table 3.5: Alternative Model Specifications

	Hazard Ratio (95% Confidence Interval)		
	Quartiles of either stress testing or catheterization	Bootstrapped Standard Errors	Shared Frailty model
Death			
Quartile 1 (Lowest)	1.00(Reference)	1.00 (Reference)	1.00 (Reference)
Quartile 2	1.06 (0.96, 1.17)	1.04 (0.94, 1.17)	1.04 (0.94, 1.15)
Quartile 3	0.99 (0.90, 1.10)	0.98 (0.89, 1.09)	0.98 (0.88, 1.08)
Quartile 4 (Highest)	1.02 (0.92, 1.13)	1.02 (0.92, 1.14)	1.02 (0.92, 1.13)
MI-related Readmission			
Quartile 1 (Lowest)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Quartile 2	0.89 (0.74, 1.06)	0.92 (0.76, 1.10)	0.92 (0.77, 1.09)
Quartile 3	1.05 (0.88, 1.25)	0.97 (0.79, 1.16)	0.97 (0.81, 1.15)
Quartile 4 (Highest)	0.91 (0.76, 1.10)	0.93 (0.76, 1.12)	0.93 (0.77, 1.11)
Repeat Revascularization			
Quartile 1 (Lowest)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Quartile 2	1.12 (1.02, 1.23)	1.13 (0.98, 1.29)	1.11 (0.99, 1.26)
Quartile 3	1.16 (1.06, 1.27)	1.11 (0.99, 1.25)	1.12 (1.00, 1.27)
Quartile 4 (Highest)	1.24 (1.13, 1.36)	1.21 (1.08, 1.34)	1.19 (1.05, 1.34)

FIGURES

Figure 3.1. Study cohort flow.

The main cohort consisted of patients receiving elective PCI for a non-acute coronary syndrome indication. The facility variation analysis cohort included elective PCI patients who had a risk factor suggesting greater benefit from surveillance stress testing and who were treated at a facility in which at least 25 patients were included in the study cohort.

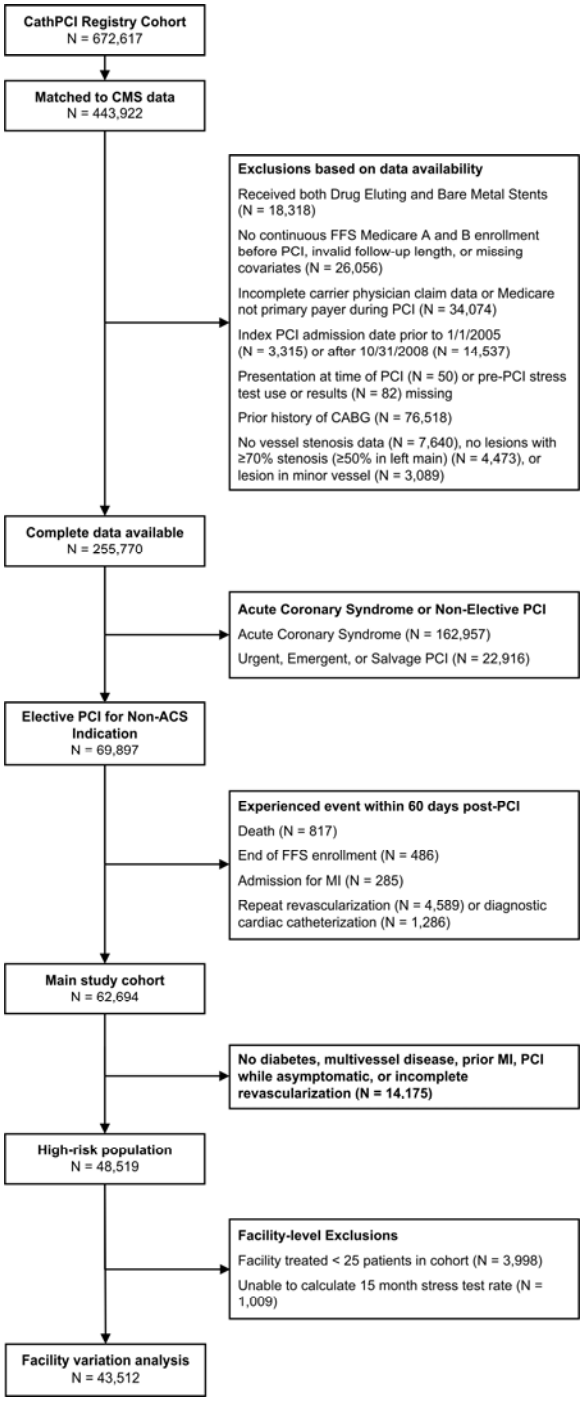


Figure 3.2. Patterns of stress testing after PCI

Panels A, C, and E illustrate the cumulative incidence of stress testing, while Panels B, D, and F illustrate kernel-smoothed estimates of the stress testing rate (the number of stress tests performed at the time point). Panels A and B illustrate the use of initial stress testing after PCI, while Panels C and D illustrate the use of a second stress after initial testing, and Panels E and F the use of a third stress test after a second test. * Competing events included death, admission for myocardial infarction, repeat coronary angiography, or repeat coronary revascularization.

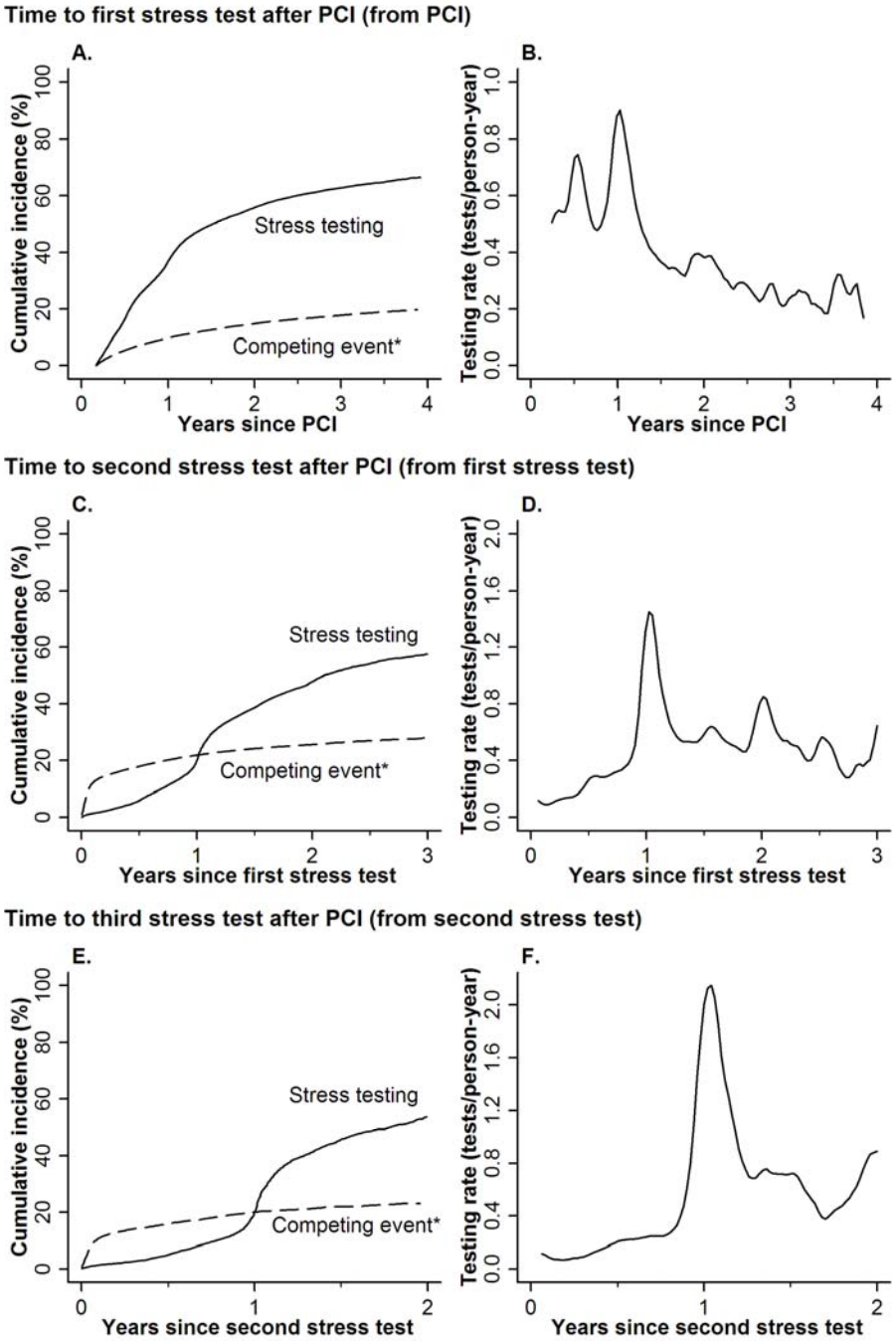


Figure 3.3. Stress testing use, stratified by patient characteristics

Bar graphs indicate unadjusted cumulative incidence of stress testing at 27 months after PCI, stratified by patient subgroups. Grey's test for difference in cumulative incidence resulted in p-values < 0.001 for all comparisons.

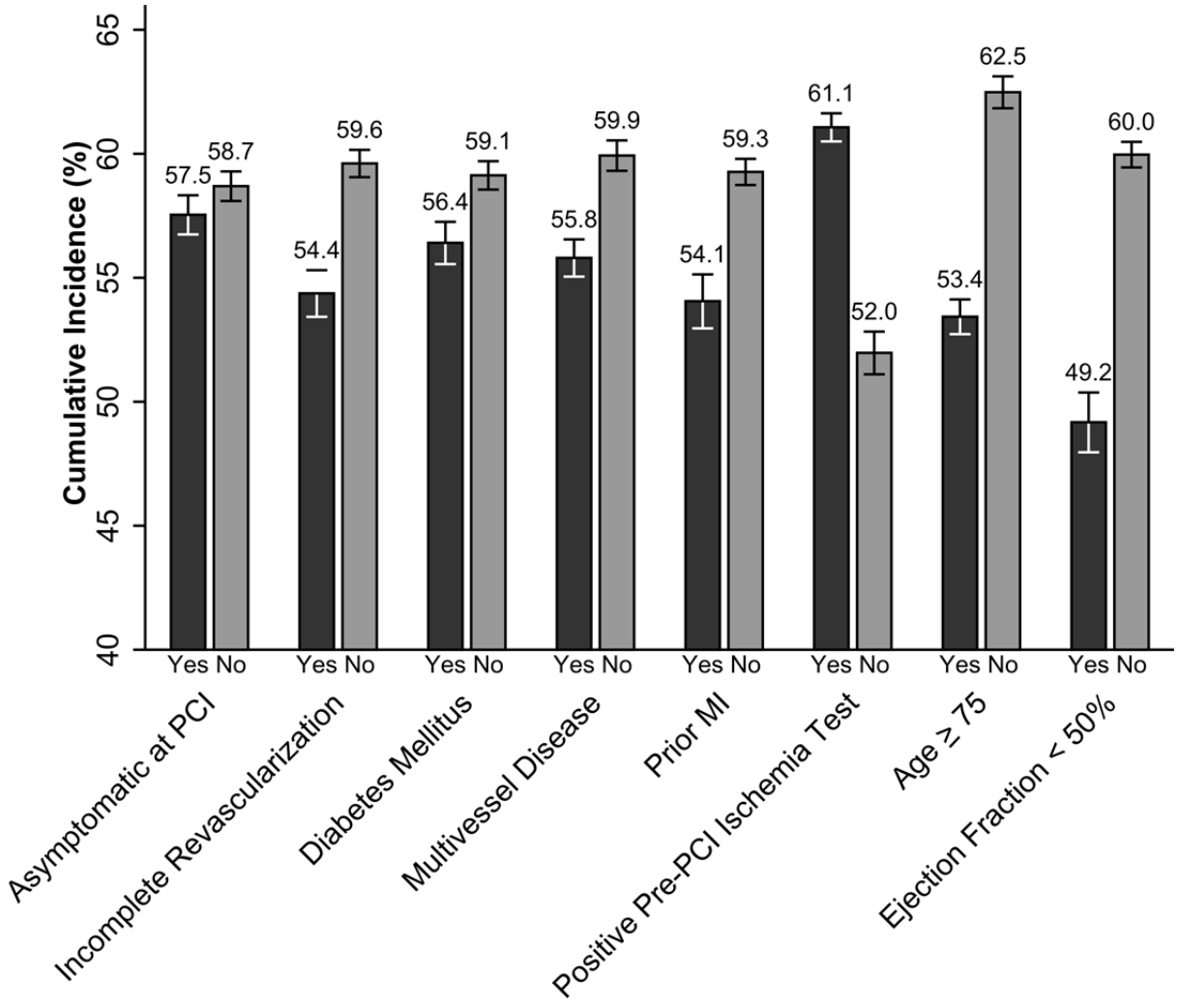


Figure 3.4. Adjusted predictors of stress testing.

Forest plot indicates adjusted predictors of receipt of any stress testing during the study period (up to 4 years). Models were fit to calculate an overall hazard ratio for each predictor. Points on the left side of the vertical line indicate an association with lower rates of stress testing, while points on the right side indicate an association with higher rates of stress testing. In addition to the main adjusted model, another model was fit to allow a time-varying effect in months 7 and 12-13 compared with the remainder of the study period. Displayed p-value is a Wald test comparing the hazard rate for the predictor in months 7, 12-13 versus remainder of study period (indirect test of differential use for surveillance testing).

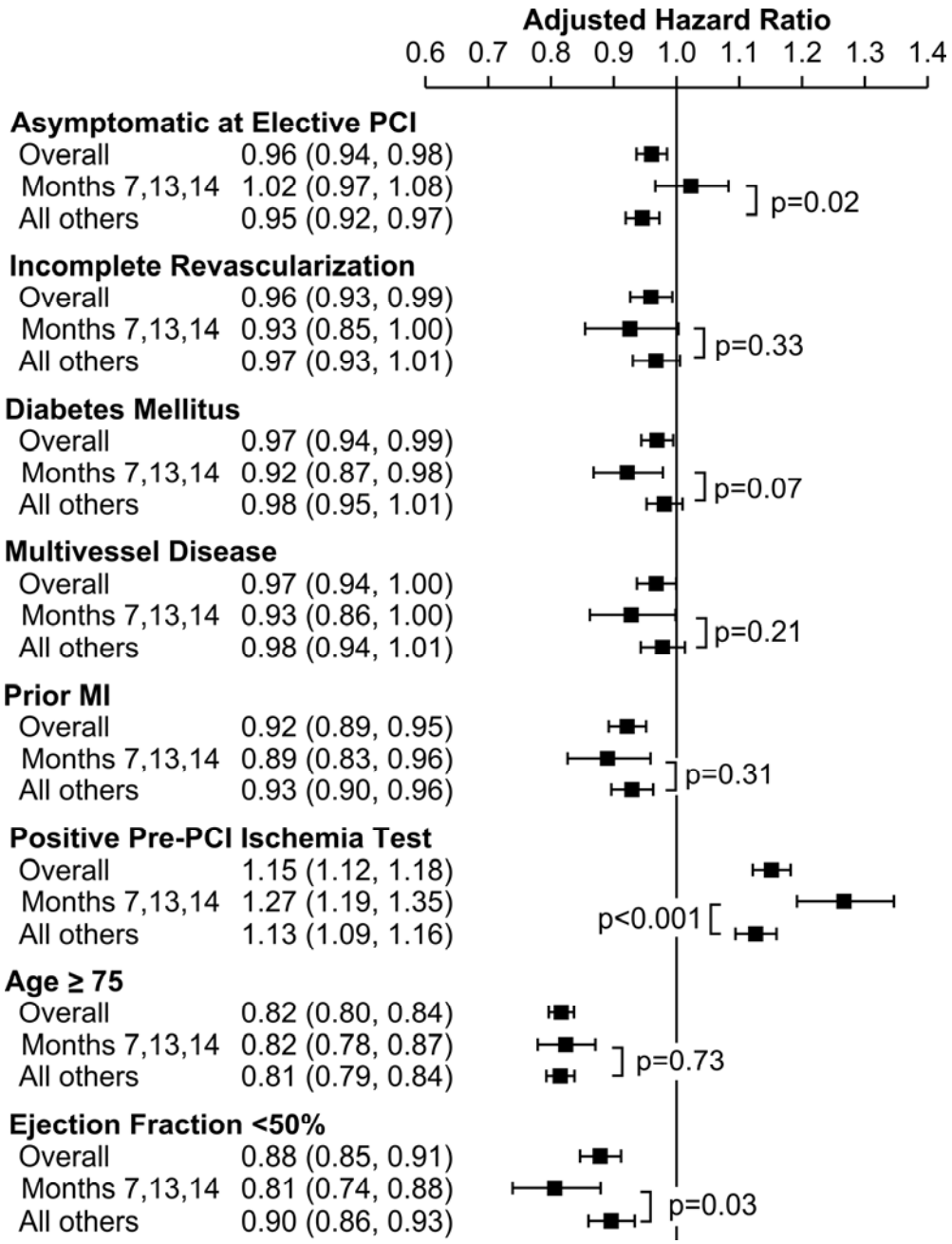


Figure 3.5. Stress testing patterns in higher-risk patients, by facility-level quartile of test use

Kernel-smoothed estimates of rate of first stress testing after percutaneous coronary intervention (PCI), stratified by facility-level quartile of stress test use within 15 months of PCI.

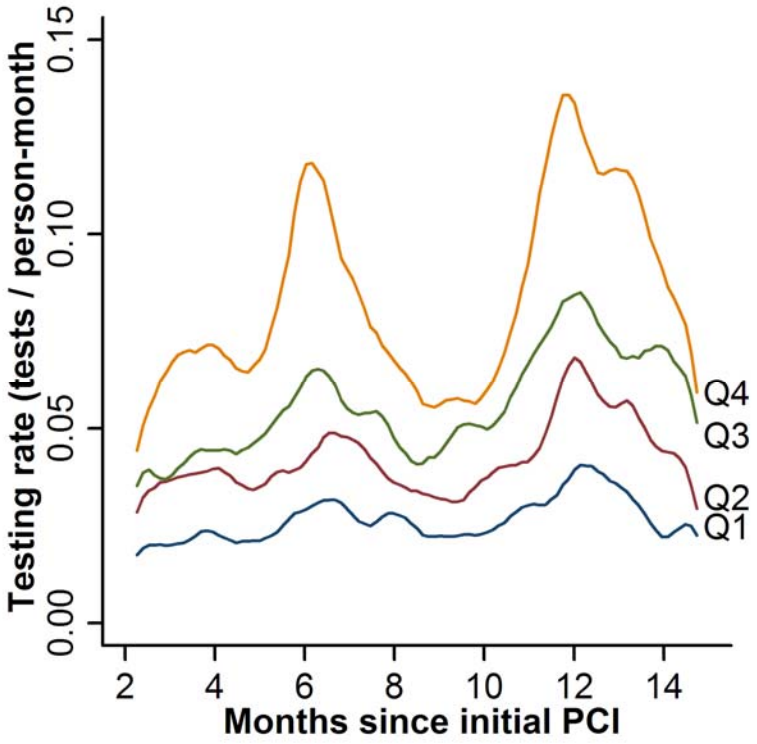
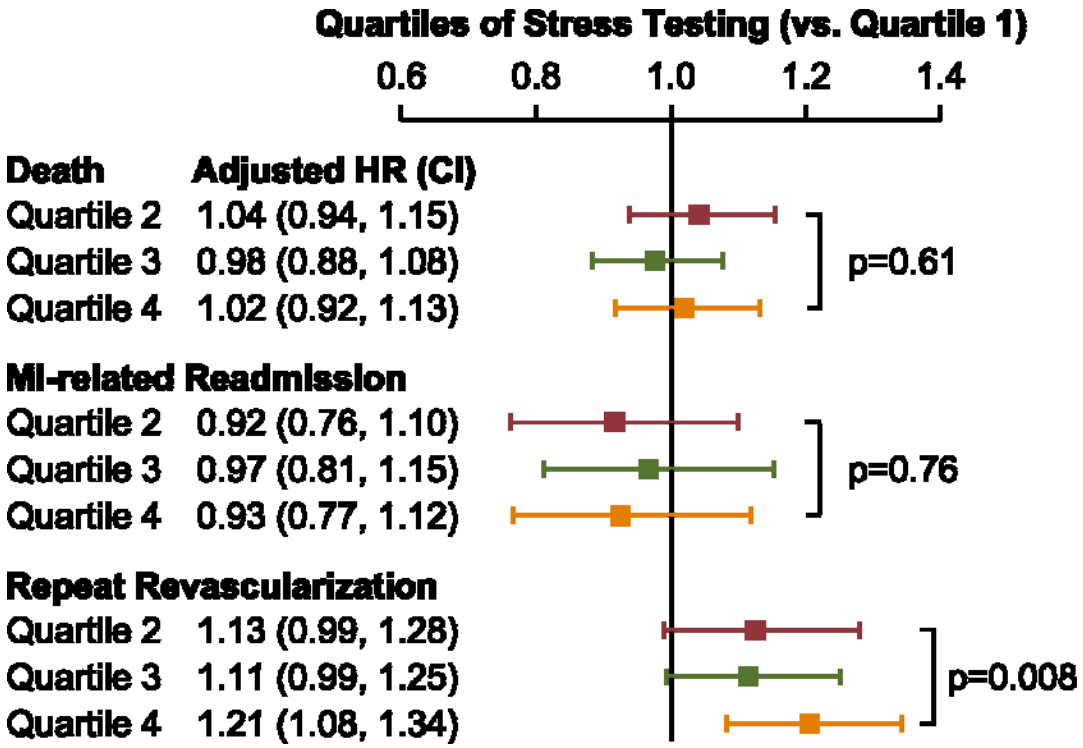


Figure 3.6. Associations between facility-level quartile of stress test use among higher-risk patients and clinical outcomes.

Forest plot indicates adjusted associations between facility-level rates of stress testing performed within 15 months of PCI and clinical outcomes. The displayed P-values are for a Wald test of joint significance of the quartile terms.



4. STUDY 2: STRESS TESTING IMAGING MODALITIES

OVERVIEW

Exercise stress testing is commonly performed following percutaneous coronary intervention (PCI) in patients with acute coronary syndromes (ACS); however, little is known about the impact of modality on patient outcomes or resource use. Our objective was to determine the impact of exercise nuclear versus exercise echocardiography testing after PCI on outcomes and resource use. This was done by doing a longitudinal observation study in the setting of medicare claims used to identify outpatient exercise stress testing with imaging within 15 months after PCI for ACS. Patients included those aged 65 and older, enrolled in fee-for-service Medicare and treated with PCI after hospitalization for ACS. Outcomes and resource use up to 4 years post-testing were compared using inverse probability weighting. We report adjusted hazard ratios (HRs) and 95% confidence intervals (CI). Among 29,279 eligible patients, 15.5% received exercise echocardiography. Echocardiography recipients had higher rates of repeat stress testing (adjusted HR: 2.60, CI: 2.19-3.10) than nuclear imaging recipients in the first 90 days after testing, but lower rates of revascularization (adjusted HR: 0.87, CI: 0.76-0.98) and catheterization (adjusted HR: 0.88, CI: 0.80-0.97). None of these differences persisted for outcomes subsequent to the first 90 days after stress testing. Rates of death and readmission for myocardial infarction rates did not differ. Total Medicare payments, including test cost, were lower initially after echocardiography, but not significantly different after 14 months post-testing. A limitation was that we lacked detailed

clinical data at the time stress testing was ordered, precluding examination of test or procedural appropriateness. In this longitudinal observational study, echocardiography recipients had fewer invasive procedures, but higher rates of repeat testing, than nuclear testing recipients. Thus, imaging cost evaluations depend on analysis duration and scope of costs considered. The comparative evaluation of imaging modalities should reflect their impact on overall process and cost efficiency.

BACKGROUND

Exercise stress testing is commonly performed following percutaneous coronary intervention (PCI) (50), most frequently using nuclear or echocardiography imaging (63). Studies have examined optimal testing approaches for patients without diagnosed coronary artery disease presenting with stable angina; however, little is known about optimal stress testing strategies in patients who have already received coronary revascularization (80, 81). Understanding the value of alternative testing modalities is critical given the costs of cardiac imaging-related services in the United States (35).

To date, most evaluations of imaging efficacy have focused on test performance parameters such as sensitivity and specificity; only recently has a research agenda been expanded to focus upon the comparative value of cardiac imaging strategies (29). Unfortunately, many imaging cost evaluations have only considered the costs associated with testing, rather than the impact of testing on resource use during or after the current episode of care. Health reform will provide incentives to move from fee-for-service reimbursement to episodic payment approaches, such as bundling, global payments, or accountable care organization; these changes will make it critical to understand the implications of test choices

on short- and long-term resource use (82). We used longitudinal administrative data to compare both clinical outcomes and resource use post-PCI among patients with acute coronary syndromes (ACS), an indication for which PCI is commonly performed.

METHODS

Data Sources and Subjects

Medicare Provider Analysis and Review (MedPAR) claims from 2003 and 2004 were obtained for all discharges with an International Classification of Diseases, 9th Edition, Clinical Modification (ICD-9-CM) diagnosis codes for myocardial infarction (410.xx), unstable angina (411.1), or angina not otherwise specified (403.9). The first such admission for each patient was considered their *index stay*. Initial exclusions restricted the sample to patients: 1) treated at a short stay facility and, 2) who did not have a diagnosis code reflecting a history of PCI, coronary artery bypass grafting (CABG), or valve replacement. Next, 2002-2006 Denominator files were obtained and additional exclusions were applied, using the Denominator data to exclude patients who were: 1) living outside of the 50 US States and Washington, DC, 2) not continuously enrolled in Fee-for-Service Medicare rather than Medicare Advantage (both exclusions to ensure complete measurement of subsequent outcomes and resource use), and 3) younger than age 66 (to allow for a full year of claims data prior to the index stay). Due to limitations on the sample size imposed by the data contractor, patients discharged after 10/14/2004 were excluded to reduce the sample size below 1,000,000 patients. **Appendix A** provides additional detail regarding initial dataset creation. For this study, the initial cohort was further restricted to individuals who were admitted for ACS (either myocardial infarction or unstable angina) and who had no

revascularization in the year preceding their index stay. For all such patients, 2003-2008 inpatient, outpatient, and carrier claims were used to determine whether patients received coronary revascularization (CABG or PCI) in the 30 days subsequent to their index stay, and, if performed, to characterize the revascularization procedures performed and subsequent outcomes. To ensure complete characterization of revascularization use, we excluded patients whose revascularization–related claims did not list Medicare as the primary payer (N = 3,170) or for which it was not possible to link either an inpatient or outpatient claim for the facility charge component of the revascularization to a Carrier claim for the professional component (N = 7,243). A small number (N = 1,895) of patients were excluded because their date of initial revascularization occurred in 2002. **Figure 4.1A** summarizes initial dataset construction, while **Table 4.1** lists the specific coding criteria employed.

Treatment Definition and Stress Test Population

The use of cardiac stress testing after PCI was identified by Healthcare Common Procedure Coding System (HCPCS) codes. Electrocardiogram stress and nuclear imaging procedures performed within one day of each other were considered a single stress nuclear event; similarly, electrocardiogram stress and echocardiographic testing performed on the same day was considered a single stress echocardiography event. Pharmacologic stress was identified using HCPCS codes; a stress test occurring on the same day (or in the case of nuclear stress testing, within one day) as a pharmacologic stress code was considered a pharmacologic stress test.

Ultimately, 162,904 PCI recipients were included for analysis **(Figure 4.1B)**. We defined a 60-day "blackout period" after each patient's index event, because diagnostic tests

during this period may be performed for cardiac rehabilitation, staging of procedures, or functional capacity assessments (50). Any use of stress testing during this period was ignored. We excluded patients who died (n=19,789), were readmitted for myocardial infarction (n=2,169), or underwent repeat revascularization or catheterization (n=15,161) during the blackout period.

Among the remaining 135,785 patients, we identified patients who received an outpatient stress test between 2 and 15 months after their index PCI event that was not preceded by diagnostic cardiac catheterization, repeat revascularization, or readmission for myocardial infarction (N = 4,921). Each patient's first eligible stress test was included in the analysis. We excluded patients: 1) receiving an inpatient stress test (based on place of service codes) in order to focus on ambulatory testing (n=5,318); 2) receiving positron emission tomography or magnetic resonance imaging (n=36) because these tests were rarely performed; 3) who were coded as having both a stress nuclear and stress echocardiography procedure on the same day (n=113); 4) receiving an ECG-only test (N = 4,135); and 5) who received a pharmacologic stress test, as pharmacologic stress testing was overwhelmingly (~98%) performed with nuclear imaging which precluded a comparison between pharmacologic nuclear and echo testing (N = 16,846).

Outcome Definitions

The use of cardiac stress testing after PCI was identified by Healthcare Common Procedure Coding System (HCPCS) codes. Electrocardiogram stress and nuclear imaging procedures performed within one day of each other were considered a single stress nuclear event; similarly, electrocardiogram stress and echocardiographic testing performed on the

same day was considered a single stress echocardiography event. Pharmacologic stress was identified using HCPCS codes; a stress test occurring on the same day (or in the case of nuclear stress testing, within one day) as a pharmacologic stress code was considered a pharmacologic stress test.

We considered seven outcomes: all-cause mortality, readmission for myocardial infarction, coronary revascularization, cardiac catheterization, additional stress testing, a composite endpoint of catheterization and additional stress testing, and total Medicare payments. The number and dates of repeat catheterizations and coronary revascularization (either PCI or coronary artery bypass grafting surgery) following stress testing were identified using ICD-9-CM procedure, HCPCS, and diagnosis-related grouping (DRG) codes. Medicare payments were obtained by measuring total payments on inpatient, outpatient, and carrier claims occurring on and/or after the date of stress testing. In addition to overall payments, we also evaluated payments stratified by whether the claim included a stress test-related line item.

Other Covariates

The presence of comorbid conditions was assessed using the Elixhauser criteria, based on diagnosis codes recorded on all claims for services performed in the year preceding the index ACS event (83). Because diagnosis codes on Carrier claims are not validated, comorbidities recorded solely on Carrier claims were only included if they were documented on two or more claims for dates of service 30 or more days apart (49). PCI facility characteristics were obtained from Medicare Provider of Services Files. Small area

socioeconomic data, measured at the ZIP code level, was obtained from the 2000 United States Census.

Statistical Analysis

We present percentages and medians with interquartile ranges (IQR) for categorical and continuous variables, respectively. Baseline patient characteristics were compared between imaging modalities using Pearson chi-square tests for categorical and Kruskal-Wallis tests for continuous variables.

A propensity score model, estimating the probability of receiving echocardiography versus nuclear imaging, was constructed using a logistic regression adjusting for patient- and facility-level characteristics. Included covariates **(Table 4.2** and **Figure 4.4)** were selected *a priori* based on known factors that may confound the relationship between imaging modality and the measured outcomes (84) including: patient demographics; ZIP code level socioeconomic characteristics, PCI procedure facility characteristics; time from PCI to stress testing; Medicare payments in the 60 days preceding stress testing; and whether the patient received medical care for comorbid conditions in the year prior to their index stay. Inverse probability weighting with stabilized weights was used to adjust for differences in the baseline characteristics of patients (85). This method offers greater precision than propensity score matching and estimates the average treatment effect of one technology versus another (analogous to the estimate from clinical trials), rather than the effect of a technology only on patients who received it (86). Since p-values can be misleading after propensity score adjustment, we evaluated the balance of covariates before and after inverse probability weighting using standardized differences (87).

All time-to-event outcomes were analyzed in a framework in which death was a competing event and end of fee-for-service claims availability was a censoring event. In addition, for the catheterization and revascularization outcomes, readmission for myocardial infarction was considered an additional competing risk; for repeat stress testing, readmission for myocardial infarction, revascularization, and catheterization were all considered competing events. Unadjusted results were portrayed using cumulative incidence curves, and bivariate tests of association performed between imaging modalities using Gray's test (75). Adjusted hazard ratios (HRs) were generated using cause-specific Cox proportional hazards models, using inverse probability weighting. To determine whether stress testing imaging modality affected short-term processes of care differently than long-term processes, we constructed models with an interaction term between imaging modality and time (a binary indicator for being within 90 days after stress testing). The 90-day threshold was specified *a priori* to distinguish "short-term" and "long-term" processes of care; however, we conducted sensitivity analyses with thresholds set at 30-day increments between 30 and 180 days post-stress testing. We also evaluated an alternative model formulation explicitly modeling the cumulative incidence of competing risks, rather than cause-specific hazards (17).

CMS payments were estimated in a partitioned framework to permit censoring caused by differential follow-up length (88). A person-period dataset was constructed using one period for the day of stress testing followed by up to 40 additional 30-day periods. Each patient's resource use was included for all time periods in which they were completely observed. Adjusted estimates of the difference in cost were constructed using weighted linear models with clustered standard errors used to account for intra-individual correlation in the person-period observations.

To address potential confounding by indication due to unobserved factors that affect test choice, we constructed facility-level variation models to compare outcomes and service use. In these models, we substituted the *percentage* of patients who received PCI at a given facility who received echocardiography for the variable indicating whether the *individual* patient received echocardiography. We hypothesized that, after adjustment, variation in site-level use of echocardiography may represent idiosyncratic small-area variation in practice patterns. To reduce measurement error, we only included patients treated in facilities where at least 50 eligible patients received an exercise stress testing with imaging. Modeling was performed using covariate-adjusted cause-specific Cox models for time to event outcomes and a partitioned estimator for total CMS payments, with standard errors clustered by PCI site to account for intra-facility correlation in observations.

Statistical analyses were conducted using SAS version 9.2 (SAS Institute, Inc., Cary, NC), R version 2.11.1 (R Foundation, Vienna, Austria), and Stata/MP version 12.1 (Statacorp, College Station, TX) with a two-sided alpha level of 0.05 pre-specified as significant. The University of North Carolina at Chapel Hill Institutional Review Board granted a waiver of the informed consent and authorization for this study.

RESULTS

Characteristics of Study Population

Of 29,279 eligible patients **(Figure 4.1b)**, 4,542 (15.5%) received exercise echocardiography. While generally similar, patients receiving exercise echocardiography had fewer comorbidities (e.g., heart failure pulmonary circulation disease, peripheral vascular disease, chronic lung disease, and diabetes mellitus) than those receiving exercise nuclear

(Table 4.2). Patients receiving exercise echocardiography were also more likely to have received a drug-eluting (rather than bare metal) coronary stent. There were no differences in the rates of the various ACS diagnoses. Echocardiography patients received PCI in smaller facilities.

Propensity Score Model

Among echocardiography recipients, propensity scores ranged from 0.02–0.51, with median 0.16 (IQR, 0.12–0.25); among nuclear imaging recipients, propensity scores ranged from ranged from 0.02–0.51, with median 0.13 (IQR, 0.10–0.17) **(Figure 4.3)**. Virtually all (99.9%) observations fell in the region of common support of the propensity score, suggesting that the analysis does not entail comparison of distinct populations. The initially selected logistic regression model, constructed without interactions or higher-order polynomial terms, produced an excellent balance of covariates as evidenced by reductions in absolute standardized differences after inverse probability weighting **(Figure 4.4)**.

Facility Variation Analysis

The 177 PCI-performing sites from which ≥50 patients received a PCI followed by an imaging-based exercise stress test during the follow-up period included 13,518 patients (46.1% of the overall sample). The proportion of patients undergoing PCI at each site who subsequently received echocardiography rather than nuclear imaging ranged from 0 to 85% (median 8.1%; IQR, 3.3–16.1%). The percentage of patients receiving echocardiography was strongly correlated across time, based on whether the patient received PCI in 2003 versus 2004 (Spearman's rho 0.73, $p<0.001$ for independence). This finding suggests the observed

facility-level variation is a durable facility characteristic rather than an artifact of the patients sampled **(Figure 4.5)**.

Outcomes and Resource Use

Follow-up data were available for a median of 1,666 days after stress testing (IQR, 1,446–1,854 days) and were virtually identical for echocardiography and nuclear testing (1,667 vs. 1,666 days, p=0.81). In unadjusted analyses, patients receiving exercise echocardiography had lower rates of repeat revascularization and catheterization after testing than did exercise nuclear patients, but higher rates of repeat stress testing and a composite of catheterization and repeat stress testing **(Figure 4.2)**. The rate of catheterization was higher for nuclear than echocardiography recipients at 90 days (14.9% vs. 12.7%) and 3 years (34.8/% vs. 32.0%); nuclear patients also had higher rates of repeat revascularization at both 90 days (9.0% vs. 7.7%) and 3 years (20.0% vs. 18.7%). Echocardiography patients had more repeat stress tests at 90 days post-testing (5.0% vs. 1.8%) and at three years (60.3% vs. 55.7%).

In adjusted analyses, overall rates of catheterization were slightly lower in exercise echocardiography recipients (adjusted HR, 0.93; 95% confidence interval [CI], 0.88–0.98), but rates of repeat stress testing (adjusted HR, 1.07; 95% CI, 1.03–1.12) were higher. Differences in overall rates of death, admission for myocardial infarction, and repeat revascularization were not statistically significant. After the first 90 days post-testing, rates of all events were equal in patients tested with echocardiography and nuclear imaging. Rates of death and admission for myocardial infarction were not significantly different in either time period. Facility variation analysis produced point estimates that were generally consistent with inverse probability weighting for the overall rates of outcomes after stress testing, with

the exception that echocardiography was associated with higher short-term mortality in the facility variation analysis. Results were consistent across a range of cut points 30-180 days following stress testing **(Table 4.5)** and when estimated using competing risk regression models rather than cause-specific hazards models **(Table 4.6)**.

Without adjustment, exercise echocardiography was associated with lower total Medicare payments than exercise nuclear on the day of stress testing (incremental difference [ID], $-497.61; CI, -506.89, -488.32); the cumulative difference between exercise echocardiography and exercise nuclear grew throughout the study period **(Figure 4.6)**. After IPW adjustment, the difference in payments on day of test was comparable to the unadjusted estimate (ID, $-499; CI, -510, -489); however, the cumulative payment difference between echocardiography and nuclear imaging decreased over time, and was no longer statistically significant at 14 months post-stress testing. Stratifying payments into those for stress testing versus all other services illustrated that stress testing-related payments were lower for exercise echocardiography recipients, but there was no significant difference in payments for all other services **(Figure 4.7)**. The $499 difference in overall payments observed on the day of initial stress test was explained by spending for stress testing-related services (difference: $491). The difference in cumulative stress testing-related payments grew smaller during the first year post-initial testing, but increased in magnitude for time points subsequent to one year.

DISCUSSION

In a national cohort of patients >65 years with ACS who had exercise stress testing with imaging after coronary stenting, patients tested with echocardiography and nuclear

imaging had similar rates of death and myocardial infarction after adjustment for baseline differences; however, the pattern of resource use post-stress testing is markedly different. Echocardiography recipients had higher rates of additional testing (catheterization and repeat stress testing combined) and higher rates of repeat stress testing, but lower rates of invasive testing and intervention, in the short-term. Moreover, costs differed depending on the time interval considered, with total CMS payments being lower among echocardiography recipients immediately after testing, but not significantly different over the long-term.

While the baseline clinical characteristics of patients tested with echocardiography versus nuclear imaging had few large differences, patients tested with echocardiography had a lower burden of risk factors than those tested with nuclear testing, including lower rates of most comorbidities. Such differences have been suggested in previous research, but have generally not been carefully examined despite their potential to skew conclusions based on Bayesian principles (89, 90). We used administrative data to adjust for potential confounding using inverse probability weights, a propensity score-based technique.

When compared with nuclear imaging, echocardiography has been demonstrated to be a less sensitive but more specific test in terms of diagnostic accuracy, as wall motion abnormalities are "further down the ischemic cascade" than the ischemia detectable by nuclear imaging (57, 91). These performance characteristics may contribute to the lower short-term use of catheterization and revascularization in post-PCI patients receiving echocardiography testing, a finding that has been previously demonstrated for patients without a history of coronary artery disease (80). The dissipation of this effect by 90 days after stress testing may indicate how quickly patients move through an episode of care for coronary artery disease.

Stress echocardiography resulted in higher short-term rates of repeat stress testing than did nuclear imaging. Several factors may explain this finding. As stress echocardiography studies are sometimes viewed as more challenging to interpret than nuclear images, higher short-term rates of repeat testing may reflect fewer definitive studies. Since physician confidence in test results is inversely correlated with rates of additional test use, our findings may also reflect lower physician comfort with echocardiography findings (92). Finally, nuclear stress testing may be viewed as having a longer "warranty period" (i.e., the period after a normal study in which the patient is viewed as being highly unlikely to have recurrent ischemia) than echocardiography (93). Since the effect attributed to echocardiography was stronger in the short-term than the long-term, an initial lack of confidence or clarity in test results (which should drive greater use of immediate repeat testing) seems a more relevant factor than differences in the warranty period.

Most outpatient stress testing post-PCI is employed either to assess symptoms suggestive of recurrent or progressive myocardial ischemia or to screen asymptomatic patients. Evidence suggests that revascularization procedures in patients with stable chest pain may improve quality of life, but generally do not improve survival or prevent myocardial infarction (94). Thus, our finding of similar rates of death and myocardial infarction (regardless of stress test imaging modality used) is consistent with the overall good prognosis of this patient population. Furthermore, documentation of similar clinical outcomes makes differences in resource use become a more important consideration.

We examined multiple contributors to costs for the two imaging stress test strategies. Stress echocardiography's lower initial cost and reduced rate of invasive procedures, but increased rate of repeat stress testing, meant that the direction of its effect on total CMS

payments relative to nuclear stress could not be predicted *a priori*. We found that stress echocardiography was associated with lower CMS payments on the day of, and up to 14 months after, initial testing. After 14 months, CMS payments were still lower for patients receiving echocardiography, but were no longer significant. The difference in test reimbursement for echocardiography versus nuclear imaging during the time period under study (2003-2005) appeared to be a stronger determinant of cost difference to payers rather than downstream service use, suggesting that equalization of reimbursement for echocardiography and nuclear testing may result in even more similar long-term costs to payers. The pattern also suggests that careful attention to decisions regarding invasive work-up shortly after stress testing may be a promising approach to controlling cost without affecting patient outcome, particularly in light of the similarities in death and myocardial infarction rates between imaging modalities.

In aggregate, our findings illustrate several complexities when evaluating cardiac imaging and using results to inform clinical practice and health policy. The differences observed in the populations chosen for each test suggest that careful attention to the risk profile of cohorts under study is needed to accurately compare testing strategies. This may be accomplished through careful statistical adjustment (as we have done) or by randomization as has been done in a handful of studies, including the recently-completed What Is the Optimal Method for Ischemia Evaluation in Women (WOMEN) Trial (2) and the ongoing PROspective Multicenter Imaging Study for Evaluation of Chest Pain (PROMISE) study (NCT01174550).

The differences in downstream testing and procedure rates but similar rates of outcomes—even after statistical adjustment for clinical and other characteristics—

demonstrate the need to examine parameters other than diagnostic and prognostic accuracy when evaluating testing strategies. In general, outcomes studies for cardiac imaging should consider the short- and long-term impact on processes of care and resource use (5). This effort may be complicated by the potential for "tradeoffs" associated with test choice: we observed that more repeat stress testing for patients receiving echocardiography was balanced by fewer invasive procedures. Furthermore, the length of follow-up is also an important consideration: we found some differences in processes of care limited to only the days immediately following stress testing (increased catheterization and revascularization with nuclear), while others persisted over time (increased repeat stress testing with echocardiography). Since the potential tradeoffs in downstream test/procedure use and temporal variation in the duration of effects, predictions regarding the relative costs of different testing strategies may be unreliable. If only the short-term is considered, exercise echocardiography might appear less expensive, while if the longer-term is considered, there may be little difference between imaging modalities. As payment models increasingly transition toward paying for episodes of care rather than on a fee-for-service basis, the length of each episode may play an important role in what are perceived to be "optimal" testing strategies.

Our study had several limitations. First, the study cohort was limited to patients >65 years and enrolled in fee-for-service Medicare. Results derived from it may not be generalizable to other populations. As with all observational studies, confounding between imaging modality and outcome may be due to factors not included in the regression model or the available data. Finally, we lacked detailed clinical data (e.g., symptoms) at the time a stress test was ordered. As a result, we cannot identify the specific indication for which the

stress test was ordered and, therefore, cannot examine the appropriateness of stress testing or of subsequent invasive procedures.

In summary, the choice of using echocardiography versus nuclear imaging in conjunction with exercise stress testing in patients who have received PCI for ACS results in a different pattern of subsequent care. Echocardiography recipients receive fewer invasive procedures in the short-term, but face increased use of repeat stress testing. Conversely, nuclear stress testing results in higher rates of downstream catheterization and repeat revascularization. These differences in post-testing patterns of care highlight that analyses of imaging value must consider not only unit cost and performance test characteristics, but also how initial testing choice affects short- and long-term outcomes and processes of care. Such considerations will be relevant for policy makers and providers in designing new reimbursement schemes for patients with ACS.

TABLES

Table 4.1: Healthcare Common Procedure Coding System and International Classification of Diseases, Ninth Revision, Clinical Modification codes used to define stress testing and outcomes.

	HCPCS, ICD-9-CM, or DRG Code	**Type of Claim**
Stress Testing		
Electrocardiogram stress test*	93015-93018	Carrier/Outpatient
Nuclear imaging	78460-78461, 78464-78465, 78472-78473, 78481, 78483	Carrier/Outpatient
Echocardiography	93350	Carrier/Outpatient
Positron Emission Tomography	78491-78492	Carrier/Outpatient
Pharmacologic stress testing	J0152 (adenosine), J1245 (dipyridamole), J1250 (dobutamine), J3490, C9399, C9244 (regadenoson)	Carrier/Outpatient
Magnetic Resonance Imaging	75552-75556 (2005-2007) 75559-75560, 75563-75564 (2008)	Carrier/Outpatient
Coronary Angiography	93508, 93539, 93540, 93545	Carrier, Outpatient
Percutaneous Coronary Intervention	HCPCS: 92980-92982, 92984, 92995, 92996, G0290, G0291, ICD-9: 36.01, 36.02, 36.05, 36.06, 36.07 DRG: 516-518, 526, 527, 555-558	Carrier, Outpatient, Inpatient
Coronary Artery Bypass Grafting	HCPCS: 33510-33514, 33516-33519, 33521-33523, 33533-33536 ICD: 36.1x DRG: 106-109, 547-550	Carrier, Outpatient, Inpatient
Acute Myocardial Infarction	410.x1 (principal diagnosis)	Inpatient

*Electrocardiogram stress and nuclear imaging procedures performed within one day of each other were considered a single stress nuclear event. Electrocardiogram stress and all other imaging performed on the same day were considered a single stress imaging event.

Table 4.2: Baseline Patient Characteristics

	Overall (N= 29,279)	Exercise Nuclear (N= 24,737)	Exercise Echo (N= 4,542)	p-value
Demographics	Median [25th-75th Percentile] or %			
Age	74.0 [70.0-78.0]	74.0 [70.0-78.0]	74.0 [70.0-78.0]	0.45
Female gender	42.5	42.6	42.3	0.71
Nonwhite race	6.2	6.1	6.8	0.07
State Medicaid buy-in	6.9	6.8	7.3	0.27
ZIP Code Characteristics				
Median household income ($10000s)	4.1 [3.4-5.3]	4.1 [3.4-5.3]	4.2 [3.4-5.4]	0.02
Percentage living in poverty	8.8 [5.2-13.8]	8.8 [5.1-13.8]	9.0 [5.5-13.8]	0.02
Percentage with college education	26.9 [19.3-39.6]	26.7 [19.1-39.2]	28.0 [20.1-41.4]	<0.001
Census region				<0.001
New England	5.6	5.6	6.1	
Middle Atlantic	14.2	14.7	11.4	
East North Central	20.2	20.6	18.5	
West North Central	10.1	9.6	12.6	
South Atlantic	21.5	23.3	11.8	
East South Central	6.4	6.9	3.7	
West South Central	8.4	8.3	8.6	
Mountain	5.1	4.4	8.6	
Pacific	8.5	6.7	18.8	
Clinical Characteristics				
ACS diagnosis				0.12
UA	45.1	45.0	45.6	
NSTEMI	23.0	23.2	21.9	
STEMI	31.9	31.7	32.5	
Multivessel PCI	19.5	19.5	19.4	0.82
Stent insertion	97.1	97.1	96.8	0.22
Any DES insertion	45.2	44.9	47.0	0.01
Any BMS insertion	53.5	53.9	51.1	<0.001
Comorbidities				
Heart failure	12.6	13.0	10.5	<0.001
Valve disease	8.1	8.1	7.7	0.31
Pulmonary circulation disease	2.0	2.0	1.6	0.04
Peripheral vascular disease	12.7	12.9	11.6	0.02

Paralysis	0.8	0.8	0.6	0.07
Neurological disease	2.5	2.5	2.2	0.25
Chronic lung disease	19.1	19.4	17.1	<0.001
Diabetes: Uncomplicated	21.7	22.1	19.7	<0.001
Diabetes: Complicated	5.3	5.5	4.2	<0.001
Hypothyroidism	12.4	12.4	12.4	0.97
Renal failure	2.8	2.8	2.6	0.33
Liver disease	0.5	0.5	0.5	0.70
Peptic ulcer	0.1	0.1	0.2	0.08
HIV/AIDS	0.0	0.0	0.0	0.94
Lymphoma	0.7	0.7	0.7	0.54
Metastatic cancer	0.6	0.6	0.6	0.70
Solid tumor	8.7	8.7	9.0	0.51
Arthritis	3.4	3.4	3.4	0.83
Liver disease	1.7	1.8	1.3	0.02
Obesity	5.8	5.9	5.0	0.02
Weight loss	0.9	0.9	0.9	0.78
Electrolyte abnormality	9.9	10.0	9.5	0.28
Blood loss anemia	1.8	1.9	1.6	0.12
Deficiency* anemia	13.0	13.2	12.0	0.03
Alcoholism	0.8	0.8	0.7	0.66
Drug use	0.1	0.1	0.1	0.43
Other psych. disease.	1.9	2.0	1.5	0.02
Depression	5.3	5.3	5.2	0.71
Hypertension	77.3	77.7	75.1	<0.001
Recent stroke	6.6	6.7	6.0	0.06
Facility Characteristics				
Hospital Ownership				0.05
Non-Profit	80.6	80.5	81.3	
For-Profit	10.1	10.3	9.1	
Government	9.3	9.2	9.6	
Major medical school affiliation	32.0	31.9	32.7	0.27
Number of CMS-authorized beds	411 [285-597]	420 [289-606]	375 [269-537]	<0.001
Pre-Stress Test Characteristics				
Medicare payments 1-60 days pre-test ($)	274 [102-687]	276 [104-691]	258 [92-667]	<0.001
Calendar time in days	408 [247-598]	402 [246-598]	430 [257-600]	<0.001

(1 = 3/1/2003)				
Time from PCI to stress test (days)	180 [112-273]	181 [113-273]	177 [107-274]	0.04
Stress test during blackout period	20.2	19.9	21.8	0.005

* Deficiencies include iron, vitamin B12, folate, and protein

Table 4.3: Propensity Score Model

	Odds Ratio [95% Confidence Interval]
Age (per 10 years)	0.951 [0.896, 1.011]
Female gender	1.043 [0.973, 1.118]
Nonwhite race	1.050 [0.911, 1.210]
Medicaid buy-in	0.983 [0.857, 1.128]
Median household income (10000s)	0.949** [0.916, 0.984]
Poverty (per 10%)	0.991 [0.926, 1.062]
College (per 10%)	1.065*** [1.029, 1.101]
Census Region	
New England	1 (reference)
Middle Atlantic	0.753*** [0.641, 0.885]
East North Central	0.940 [0.804, 1.098]
West North Central	1.349*** [1.143, 1.593]
South Atlantic	0.523*** [0.443, 0.618]
East South Central	0.575*** [0.463, 0.715]
West South Central	1.115 [0.928, 1.339]
Mountain	1.934*** [1.610, 2.323]
Pacific	2.827*** [2.399, 3.331]
ACS Diagnosis	
Unstable Angina	1 (reference)
Non-ST Elevation Myocardial Infarction	0.941 [0.865, 1.024]
ST Elevation Myocardial Infarction	0.954 [0.883, 1.030]
Multivessel procedure	0.982 [0.903, 1.067]
PCI w/ stent insertion	1.004 [0.788, 1.279]
PCI w/ any DES insertion	0.913 [0.776, 1.075]
PCI w/ any BMS insertion	0.874 [0.745, 1.025]
Comorbidities	
Heart failure	0.840** [0.751, 0.939]
Valve disease	1.043 [0.921, 1.181]
Pulmonary circulation dz.	0.883 [0.683, 1.141]
Peripheral vascular dz.	1.000 [0.902, 1.109]
Paralysis	0.772 [0.498, 1.198]
Neurological dz.	0.948 [0.761, 1.182]
Chronic lung dz.	0.927 [0.849, 1.012]
Diabetes: Uncomplicated	0.890** [0.819, 0.968]
Diabetes: Complicated	0.790** [0.670, 0.930]
Hypothyroidism	1.015 [0.918, 1.123]
Renal failure	1.057 [0.855, 1.306]
Liver disease	1.191 [0.759, 1.871]

Peptic ulcer	1.639 [0.777, 3.457]
HIV/AIDS	1.496 [0.173, 12.94]
Lymphoma	0.919 [0.617, 1.369]
Metastatic cancer	1.021 [0.670, 1.557]
Solid tumor	1.088 [0.968, 1.224]
Arthritis	1.016 [0.848, 1.217]
Liver disease	0.816 [0.617, 1.079]
Obesity	0.863 [0.744, 1.001]
Weight loss	1.026 [0.720, 1.463]
Electrolyte abnormality	1.056 [0.942, 1.185]
Blood loss	0.925 [0.713, 1.200]
Deficiency anemia	0.971 [0.874, 1.080]
Alcoholism	0.918 [0.626, 1.347]
Drug use	0.772 [0.262, 2.269]
Non-depression psych. dz.	0.795 [0.607, 1.040]
Depression	1.064 [0.915, 1.237]
Hypertension	0.938 [0.867, 1.015]
Recent stroke	0.968 [0.839, 1.116]
Hospital Type	
Non-profit	1 (reference)
For-profit	0.785*** [0.696, 0.884]
Government	1.129* [1.005, 1.269]
Major medical school affiliation	1.208*** [1.115, 1.309]
Number of CMS-authorized beds (per 100)	0.966*** [0.953, 0.980]
Medicare payments during 60 days pre-ST (per 100)	0.999* [0.997, 1.000]
Calendar time (1 = 3/1/2003) (per 100 days)	1.029** [1.007, 1.051]
Time from PCI to stress test (per 100 days)	0.956* [0.920, 0.994]
Received stress test during blackout period	1.087* [1.001, 1.181]

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table 4.4: Unadjusted and Adjusted Cause-Specific Hazards Models

	Unadjusted Model	IPW Model	Facility-Level Variation Model
	Entire Study Period		
Death	0.88 (0.81, 0.96)	0.99 (0.90, 1.09)	0.80 (0.57, 1.12)
MI	0.96 (0.84, 1.09)	1.01 (0.87, 1.17)	0.93 (0.55, 1.58)
Revascularization	0.93 (0.87, 0.99)	0.94 (0.87, 1.01)	0.89 (0.62, 1.27)
Catheterization	0.91 (0.87, 0.96)	0.93 (0.88, 0.98)	0.84 (0.62, 1.15)
Second stress test	1.12 (1.07, 1.16)	1.07 (1.03, 1.12)	1.28 (0.98, 1.66)
Catheterization or second stress test	1.04 (1.01, 1.08)	1.02 (0.98, 1.06)	1.11 (0.89, 1.39)
	Days 1-90 after Stress Testing		
Death	1.04 (0.68, 1.59)	1.18 (0.75, 1.87)	3.35 (1.06, 10.6)
MI	0.70 (0.45, 1.10)	0.61 (0.37, 1.02)	0.49 (0.07, 3.66)
Revascularization	0.84 (0.75, 0.94)	0.87 (0.76, 0.98)	0.82 (0.48, 1.38)
Catheterization	0.85 (0.78, 0.93)	0.88 (0.80, 0.97)	0.80 (0.51, 1.26)
Second stress test	2.78 (2.37, 3.26)	2.60 (2.19, 3.10)	2.83 (1.23, 6.54)
Catheterization or second stress test	1.04 (0.96, 1.13)	1.04 (0.96, 1.13)	1.07 (0.64, 1.77)
	Days 91+ after Stress Testing		
Death	0.87 (0.80, 0.96)	0.98 (0.89, 1.08)	0.74 (0.53, 1.04)
MI	0.99 (0.86, 1.13)	1.06 (0.91, 1.24)	0.99 (0.58, 1.68)
Revascularization	0.98 (0.90, 1.07)	0.99 (0.90, 1.08)	0.93 (0.65, 1.33)
Catheterization	0.95 (0.89, 1.01)	0.96 (0.89, 1.03)	0.86 (0.62, 1.19)
Second stress test	1.06 (1.02, 1.11)	1.03 (0.98, 1.07)	1.24 (0.95, 1.60)
Catheterization or second stress test	1.04 (1.00, 1.08)	1.01 (0.97, 1.06)	1.12 (0.88, 1.42)

All abbreviations can be found in Table 4.1.

Table 4.5: Robustness of inverse probability weighted model results to threshold value for time-varying effect

	Hazard Ratio (CI)				
	Death	**MI**	**Revasc**	**Cath**	**Second Stress Test**
Overall	1.11 (0.95, 1.30)	1.02 (0.85, 1.23)	0.94 (0.87, 1.02)	0.94 (0.89, 1.00)	1.22 (1.16, 1.28)
30-day cutpoint					
≤ 30 days	0.83 (0.33, 2.11)	1.10 (0.57, 2.14)	0.88 (0.76, 1.02)	0.89 (0.79, 0.99)	3.62 (2.74, 4.77)
> 30 days	0.99 (0.90, 1.09)	1.01 (0.87, 1.17)	0.96 (0.89, 1.05)	0.94 (0.88, 1.01)	1.05 (1.00, 1.10)
60-day cutpoint					
≤ 60 days	1.06 (0.55, 2.05)	0.68 (0.38, 1.23)	0.86 (0.76, 0.98)	0.86 (0.78, 0.95)	3.22 (2.60, 3.99)
> 60 days	0.99 (0.89, 1.09)	1.04 (0.89, 1.21)	0.98 (0.90, 1.08)	0.96 (0.90, 1.03)	1.03 (0.99, 1.08)
90-day cutpoint					
≤ 90 days	1.18 (0.75, 1.87)	0.61 (0.37, 1.02)	0.87 (0.76, 0.98)	0.88 (0.80, 0.97)	2.60 (2.19, 3.10)
> 90 days	0.98 (0.89, 1.08)	1.06 (0.91, 1.24)	0.99 (0.90, 1.08)	0.96 (0.89, 1.03)	1.03 (0.98, 1.07)
120-day cutpoint					
≤ 120 days	1.32 (0.90, 1.94)	0.84 (0.53, 1.32)	0.86 (0.77, 0.97)	0.87 (0.79, 0.95)	2.27 (1.97, 2.61)
> 120 days	0.97 (0.88, 1.07)	1.04 (0.89, 1.21)	0.99 (0.91, 1.09)	0.97 (0.90, 1.04)	1.01 (0.96, 1.06)
150-day cutpoint					
≤ 150 days	1.18 (0.83, 1.68)	0.76 (0.50, 1.17)	0.85 (0.76, 0.96)	0.87 (0.79, 0.95)	2.00 (1.76, 2.28)
> 150 days	0.98 (0.88, 1.08)	1.06 (0.90, 1.24)	1.01 (0.92, 1.11)	0.97 (0.90, 1.04)	1.01 (0.96, 1.06)
180-day cutpoint					
≤ 180 days	1.06 (0.77, 1.48)	0.80 (0.54, 1.18)	0.84 (0.75, 0.95)	0.86 (0.78, 0.94)	1.90 (1.69, 2.12)
> 180 days	0.98 (0.89, 1.09)	1.06 (0.90, 1.24)	1.02 (0.93, 1.13)	0.98 (0.91, 1.06)	0.99 (0.95, 1.04)

Cath = catheterization; CI = confidence interval; MI = myocardial infarction; Revasc = revascularization

Table 4.6: Inverse Probability Weighted Competing Risk Proportional Hazards Models

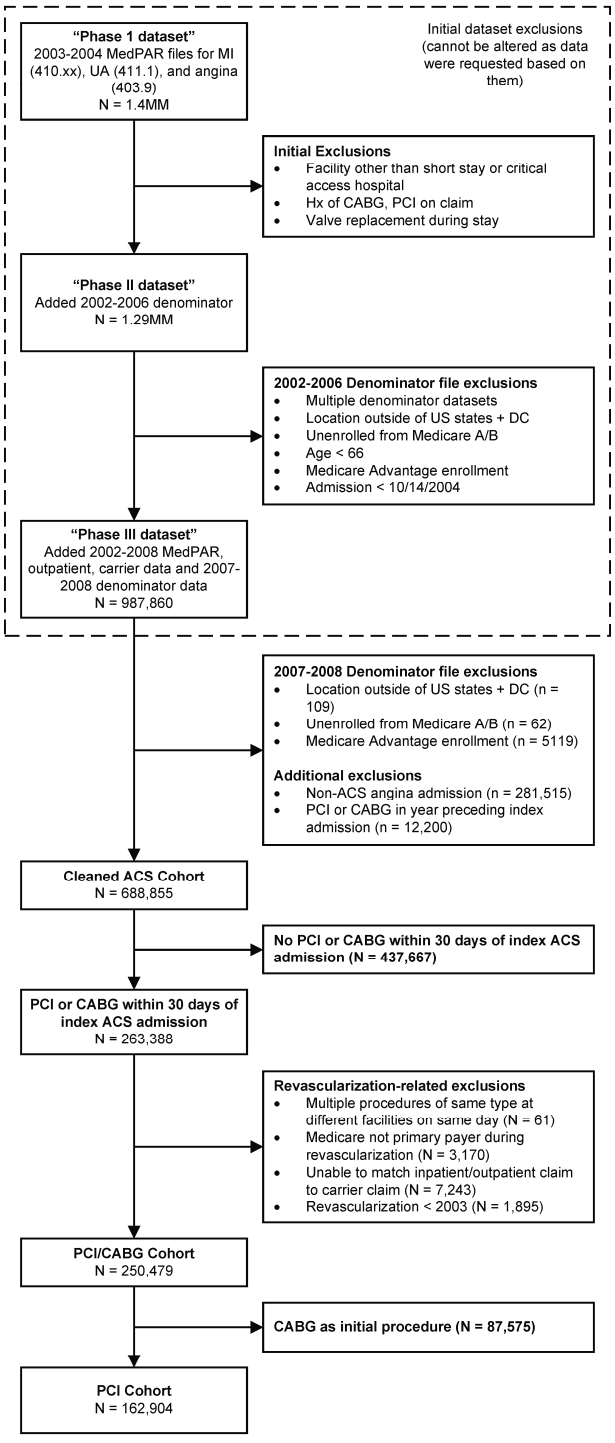
	Hazard Ratio (CI)			
	MI	**Revasc**	**Cath**	**Second Stress Test**
Overall	1.01 (0.87, 1.17)	0.94 (0.88, 1.02)	0.93 (0.88, 0.98)	1.13 (1.08, 1.18)
90-day cutpoints				
≤ 90 days	0.61 (0.37, 1.02)	0.87 (0.76, 0.98)	0.88 (0.80, 0.97)	2.64 (2.22, 3.15)
> 90 days	1.06 (0.91, 1.24)	0.99 (0.91, 1.09)	0.96 (0.90, 1.03)	1.08 (1.03, 1.13)

All abbreviations can be found in Appendix Table 1.

FIGURES

Figure 4.1A-B: Cohort Selection Diagram

Figure illustrates process by which study cohort was identified. Included patients were those receiving an outpatient exercise nuclear or echocardiography stress test not preceded by another cardiac event.



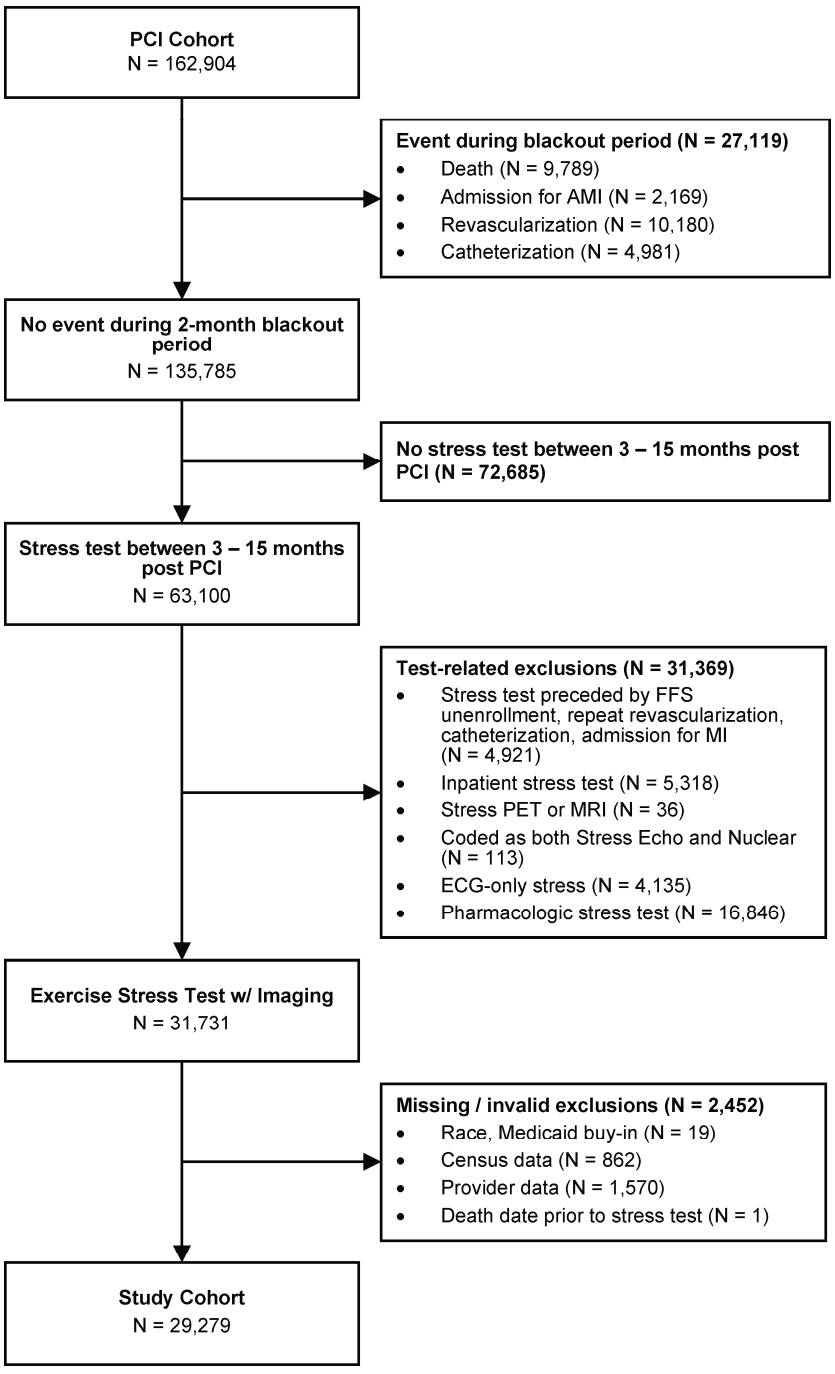

PCI Cohort
N = 162,904
Event during blackout period (N = 27,119)
Death (N = 9,789)
Admission for AMI (N = 2,169)
Revascularization (N = 10,180)
Catheterization (N = 4,981)
No event during 2-month blackout period
N = 135,785
No stress test between 3 – 15 months post PCI (N = 72,685)
Stress test between 3 – 15 months post PCI
N = 63,100
Test-related exclusions (N = 31,369)
Stress test preceded by FFS unenrollment, repeat revascularization, catheterization, admission for MI (N = 4,921)
Inpatient stress test (N = 5,318)
Stress PET or MRI (N = 36)
Coded as both Stress Echo and Nuclear (N = 113)
ECG-only stress (N = 4,135)
Pharmacologic stress test (N = 16,846)
Exercise Stress Test w/ Imaging
N = 31,731
Missing / invalid exclusions (N = 2,452)
Race, Medicaid buy-in (N = 19)
Census data (N = 862)
Provider data (N = 1,570)
Death date prior to stress test (N = 1)
Study Cohort
N = 29,279

Figure 4.2: Unadjusted Outcomes by Stress Test Imaging Modality

Curves illustrate unadjusted cumulative incidence of outcomes, based on time since stress testing. P-values are for comparison of nuclear versus echocardiography using Gray's test.

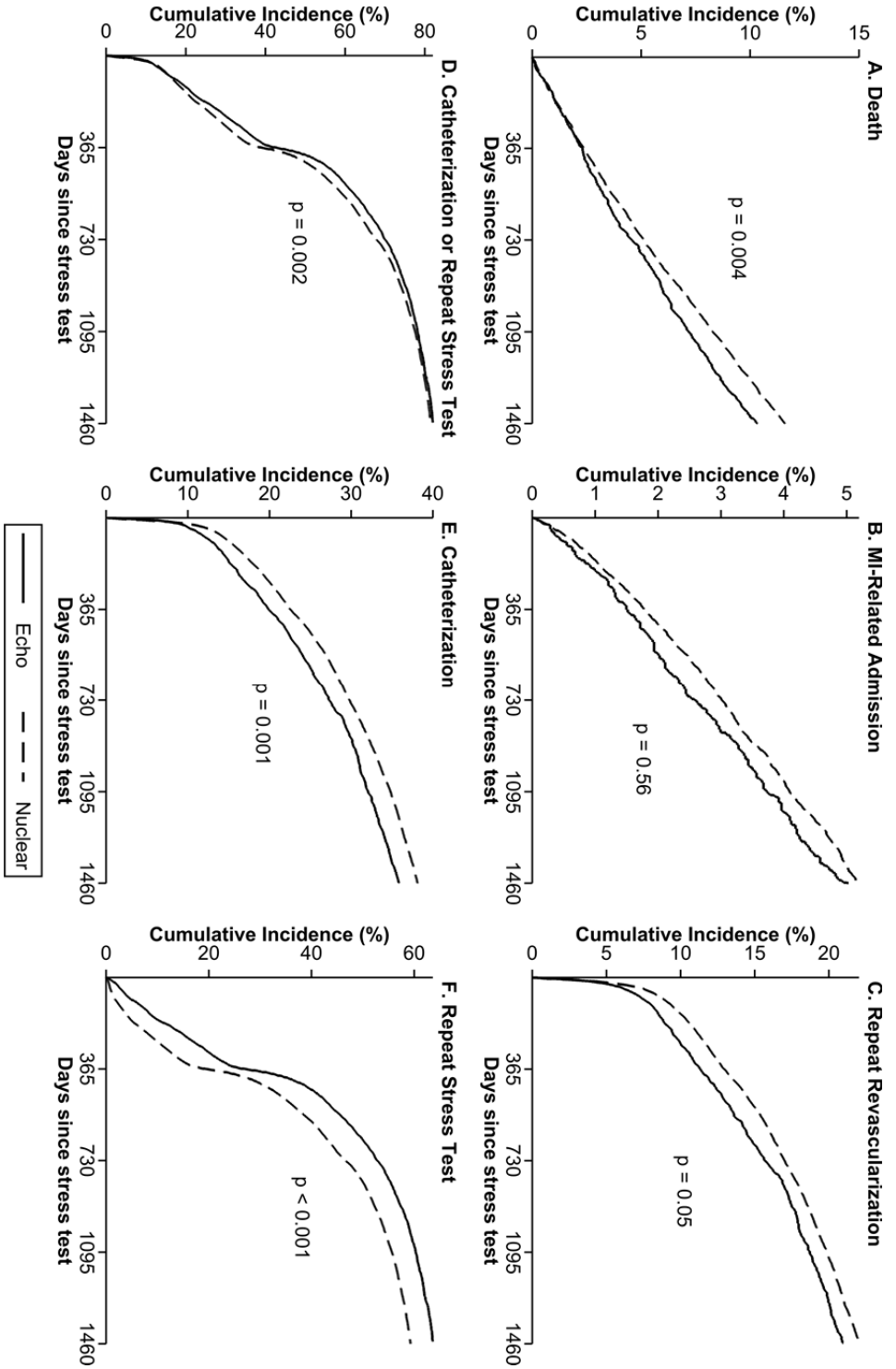


Figure 4.3: Distribution of propensity scores

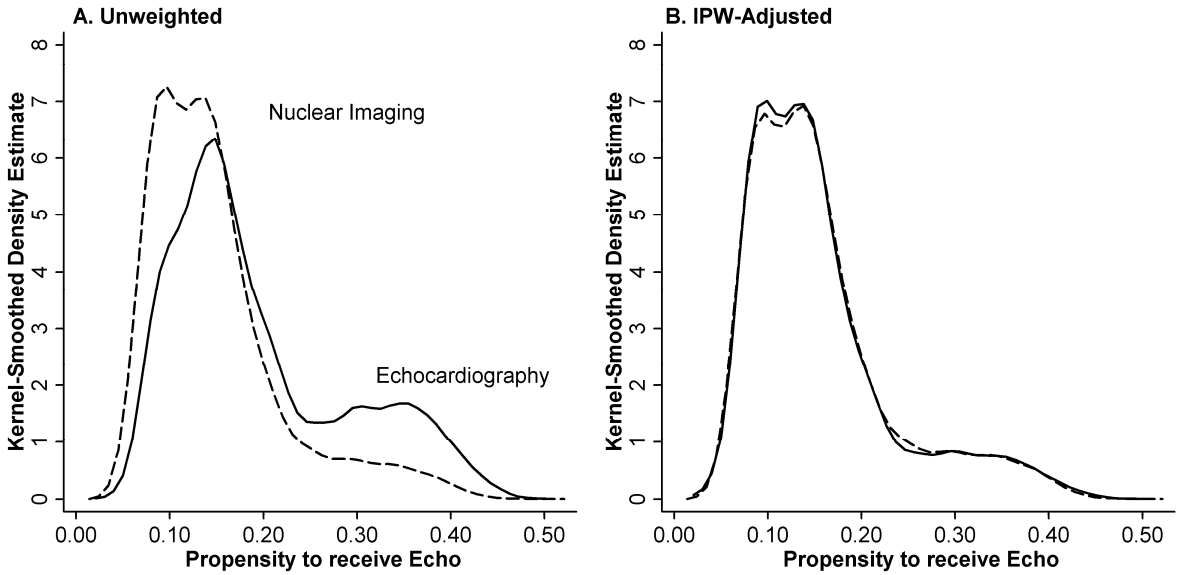


Figure 4.4: Impact of inverse probability weighting on covariate balance

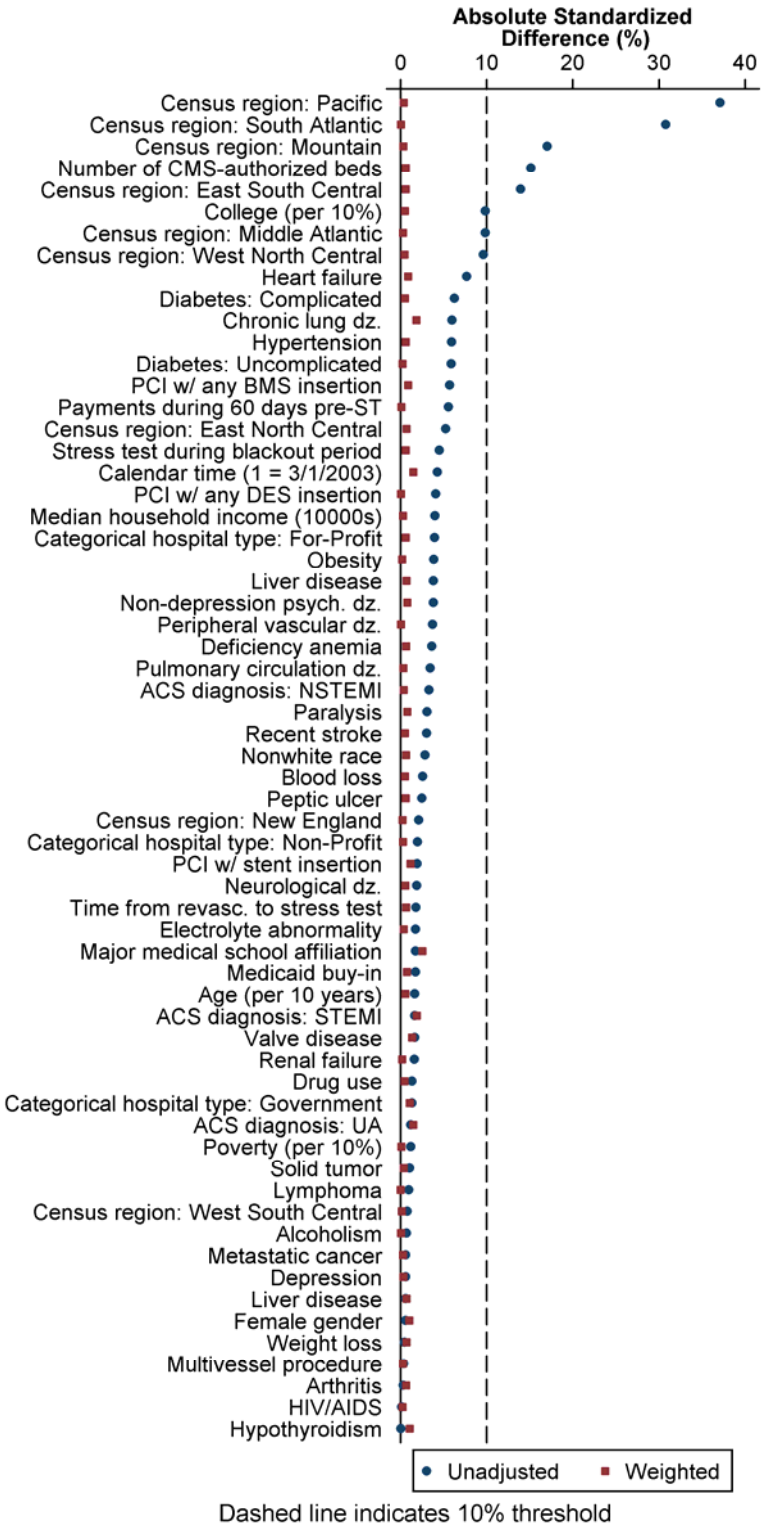


Figure 4.5: Temporal stability of facility-level imaging modality selection

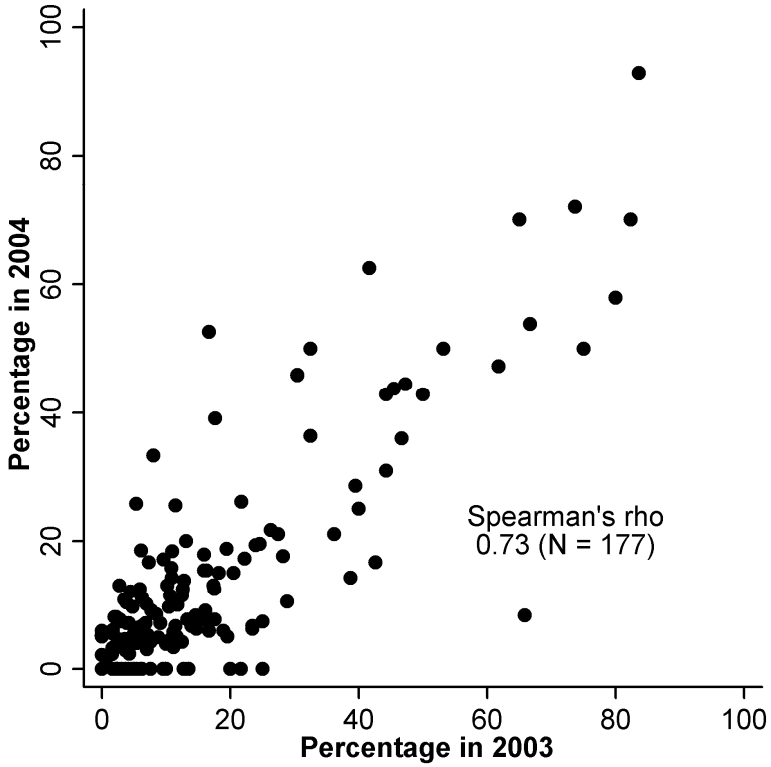


Figure 4.6: Unadjusted and Adjusted Incremental Total Medicare Payments

The top row illustrates per-period incremental costs, while the bottom row illustrates cumulative incremental costs up to the period in question. Dotted lines in the bottom row indicate incremental difference in cost accruing during the day of initial stress testing.

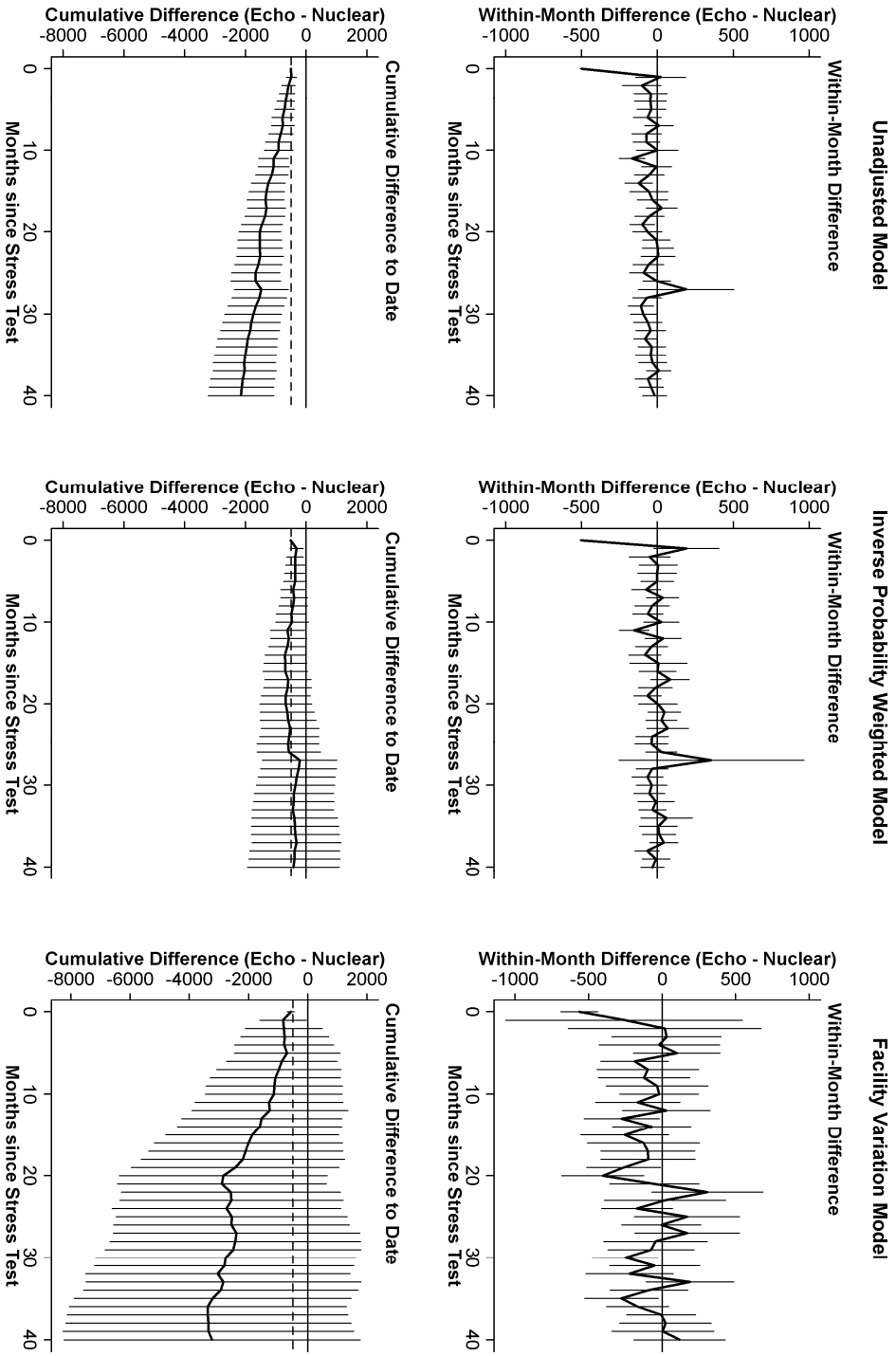
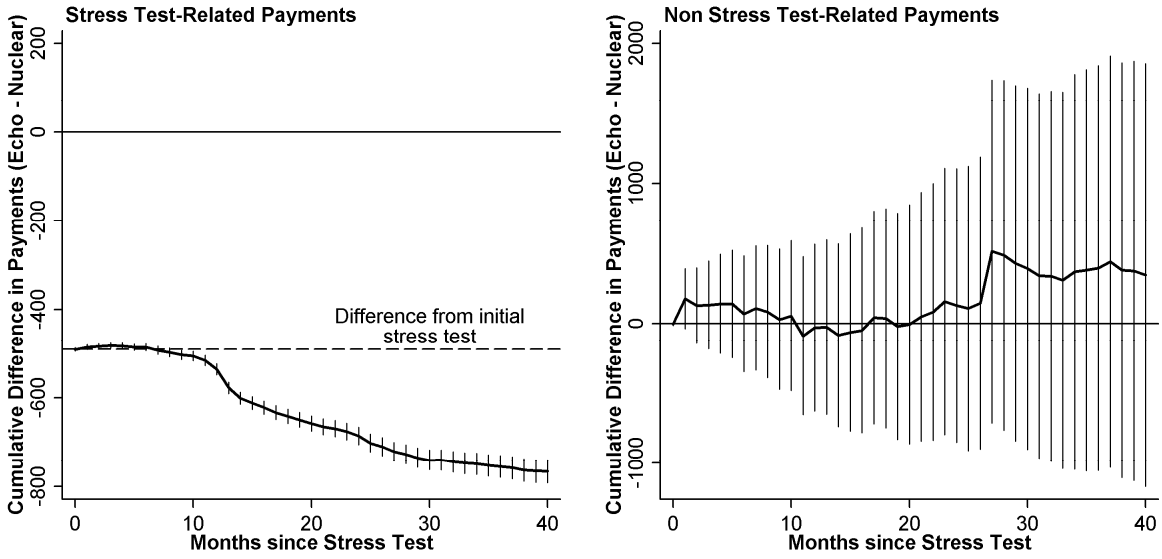


Figure 4.7: Inverse Propensity Weighted (IPW)-Adjusted Cumulative Incremental Costs

Comparing echocardiography to nuclear imaging with costs stratified by stress test-related versus all other.



5. INSTRUMENTAL VARIABLES FOR RELEVANT ESTIMATES

OVERVIEW

The use of observational methods for cardiovascular comparative effectiveness research (CER) is an increasingly attractive alternative to randomized trials. However, endogeneity (confounding) of treatment choice is a major methodological limitation. Conventional methods such as propensity score and outcome modeling can adjust for endogeneity due to factors observed in the available data, but not unobserved factors. In contrast, instrumental variable methods allow for the consistent estimation of treatment effects from observational data even when endogeneity is due to unobserved factors. Unfortunately, instrumental variable methods typically estimate a treatment effect that only applies to an unidentifiable "marginal" population of patients, which complicates the interpretation of results for clinical and policy decisions. Economists have pioneered a new method – local instrumental variables – which allows for the estimation of treatment effects that are more directly clinically and policy relevant, such as the average treatment effect. In this study, we illustrate the value of these new methods and describe how they can be applied to CER using the example of drug-eluting coronary stents. We find evidence that drug-eluting coronary stents are safe and effective in the patients receiving them, but that significant heterogeneity in treatment response exists such that patients who did not receive DES would be predicted to have a more adverse risk-benefit profile.

INTRODUCTION

Observational data in outcomes research

Due to limitations of randomized trials, the cardiovascular community has demonstrated considerable enthusiasm for comparative effectiveness research (CER) employing observational data (95, 96). Observational data allow measurement of treatment effects in "real world" clinical practice. CER using observational data is less expensive and facilitates important comparisons that would be impractical using experimental methods. The availability of powerful computing resources paired with the ever-increasing accrual of data through administrative, registry, and electronic medical records mean that observational methods will continue to be an attractive option for conducting cardiovascular CER.

Instrumental variable methods are a valuable tool for observational CER

Unlike trials, treatments are not assigned at random in observational data; thus, the potential for bias due to endogenous treatment choice (*confounding by indication*) is a persistent methodological challenge (97). Endogenous treatment choice is present when factors that affect treatment choice also affect patient outcomes through pathways other than treatment choice. The most common method of addressing endogenous treatment choice has been to identify factors affecting both treatment choice and outcome (i.e., the confounding factors) and to adjust for them with multivariate regression modeling. This adjustment can be performed through modeling of either treatment choice or outcome; when treatment choice is modeled, the approach is considered a *propensity score* method. There are circumstances in which modeling either treatment choice or outcome is clearly preferable (98). One such circumstance is the case of a common treatment but rare outcome. In this circumstance,

outcome modeling may be limited by the poor asymptotic properties of non-linear estimators when used to estimate a high number of covariates relative to number of outcomes (99). There are also methods that attempt to model both treatment choice and outcome to achieve a "doubly robust" estimator which is robust to misspecification of the treatment equation or outcome equation (but not both) (100). Despite extensive debate in the literature concerning theoretical advantages and disadvantages of each method, empirically, the methods generally produce very similar results (101). Importantly, both conventional multivariate modeling and propensity score methods share a common and significant limitation: while they can adjust for bias caused by endogenous treatment selection due to factors that are observed by the analyst (or highly correlated with these factors), they cannot adjust for bias due to factors that are either unobserved or not perfectly correlated with observed factors. Incomplete adjustment for endogenous treatment choice can lead to severely biased estimates of treatment effects.

In contrast to conventional modeling approaches, instrumental variable (IV) techniques allow measurement of treatment effects that are adjusted for endogenous treatment choice, even if the confounding factors are not completely observed. This benefit is achieved by identifying *instruments*, which are variables that affect the patient's treatment choice but have no expected effect on outcome through a mechanism other than treatment choice. In the language of econometrics, the instrument's effect on treatment choice must be plausibly exogenous; in the language of statistics and epidemiology, it must be unconfounded.[2] The seminal application of IV methods in cardiovascular medicine was an

[2] Formally, IVs need only be exogenous after conditioning on observed covariates. The formal presentation of IV in the Model section clarifies this distinction.

effort to identify the effects of early invasive treatment versus thrombolysis for myocardial infarction, an area in which conventional methods were thought to provide inadequate adjustment for underlying differences in patient risk (102). The difficulty of identifying valid instruments for specific clinical situations has limited the broad application of IVs, as IV models estimated with poor quality instruments can, like incompletely adjusted conventional methods, result in severely biased estimates.[3] However, recent successes in leveraging small area geographic variation and idiosyncrasies in provider preference as instruments have made performing such analyses increasingly tractable (103, 104).

Treatment effect heterogeneity limits existing instrumental variable methods

The benefits of IV analysis for addressing endogeneity are well established, and there has been considerable progress made in identifying valid IVs for important questions in cardiovascular CER. However, a notable and persistent limitation of IV analyses is that they provide results that are challenging interpret for use in clinical practice. This is particularly true when individual patients respond differently to a treatment, a phenomenon known as *treatment effect heterogeneity*.

Most IV analyses estimate a *local average treatment effect* (LATE), which is the average effect of the treatment on a "marginal population" of patients whose treatment choice is affected by the IV (105). Unfortunately, it is impossible to identify precisely which patients comprise this marginal population (106), and except in two special cases discussed

---

[3] One notable limitation of IV methods is the ease with which, through selection of an invalid IV, one can generate results that are more biased than an unadjusted, naïve comparison of outcomes (i.e., results that are worse than would be attained with a t-test).

later, the LATE will differ from treatment effects for other patients. There are circumstances in which the LATE estimated by IV is a policy-relevant *estimand* (result of estimation). For instance, investigators used the sharp decline in drug-eluting coronary stent (DES) utilization in late 2006 as an IV to determine how the more selective use of DES in 2007 affected patient outcome when compared with the permissive use in early 2006 (107).[4] In this case, the specific question being asked was whether those "marginal patients" who would have received DES in 2006, but not in 2007, experienced a change in outcome should providers adopt practice patterns reminiscent of 2006 or 2007 going forward. Similarly, if a government assistance program changes its eligibility criteria from 100% of the Federal Poverty Line to 133%, an IV analysis using the changed threshold as an instrument would only estimate to the effect of the government assistance program on the population between 101-133% of the poverty line. However, investigators conducting CER are often interested in questions that correspond more directly with other *estimands*:

1. *Average Treatment Effect (ATE):* The ATE is the expected difference in outcome caused by the treatment, averaged across the entire population that is considered potentially eligible to receive the treatment. Alternatively, it can be viewed as the effect of applying a treatment across an entire population of patients, such as would occur if a treatment was used in all eligible patients with a particular condition immediately after regulatory approval.

2. *Effect of Treatment on the Treated (TT):* The TT is also an average treatment effect, but it is the expected difference in outcomes with and without treatment for only those patients actually receiving treatment in clinical practice. This estimand is particularly useful

---

[4] I examined and ultimately rejected using a similar temporal instrument in this paper. It did not work because at the same time that the percentage of PCI patients receiving DES fell during this time period, the volume of PCI cases also declined.

for determining whether the treatment, as currently used, improves patient outcome. As such, it is the natural estimand for decisions such as whether a treatment should remain covered by insurance plans or remain approved by regulators for a given indication.

3. *Effect of Treatment on the Untreated (TUT):* The TUT is analogous to the TT, but it is instead the average treatment effect among those patients who do *not* currently receive the treatment. It can also be viewed as the predicted impact of expanding treatment to previously-untreated segments of the patient population. Consequently, the TUT is a natural estimand for deciding what treatments should be considered mandatory or defined as a measure of quality care.

When treatment effect heterogeneity exists, and patients or their providers select treatments based on the factors associated with the treatment effect heterogeneity, the LATE estimand produced by conventional IV is not a consistent estimate of the ATE, TT, or TUT. **Figure 5.1** summarizes the situations in which ATE, TT, and TUT differ from LATE. There has been a long-standing emphasis on estimation of the ATE in the clinical and policy literature (108) perhaps because the ATE is the estimand that most closely approximates those obtained from randomized trials.[5] As such, it is particularly appropriate for an initial demonstration of a treatment's safety and efficacy, as is done in Phase II/III trials. Because randomized trials by definition do not allow patients to choose treatments,[6] estimation of

---

[5] Studies analyzed on an intention to treat basis (i.e., those in which patients do not necessarily receive the assigned treatment) do not formally identify the ATE but rather the "average effect of being assigned to receive treatment".

[6] Many randomized trials permit treatment refusal or crossover due to well-accepted ethical limitations concerning how subject behavior can be regulated in research settings; when these are permitted, the results are not strictly an ATE. However, even when such behavior is allowed, the resultant effect estimates are also still not equivalent to TT / TUT in non-

meaningful TT and TUT parameters is only possible in observational settings.[7] It is not uncommon, however, for clinical or policy questions to match more naturally to the TT or TUT than the ATE. For example, the question of whether an existing treatment should remain available to patients is most appropriately answered by considering the treatment's effect only on those patients currently receiving it (TT). Doing so recognizes that patients and providers may selectively treat patients with the most favorable risk/benefit profile to receive treatment. Conversely, evidence that the average response to a treatment is positive is not definitive evidence that the treatment ought to be provided to all patients with the condition. Estimation of the TUT would provide stronger evidence about whether changing practice patterns to expand use of the treatment would improve outcomes.

Local instrumental variable methods address treatment effect heterogeneity

Economists have developed a method known as *local instrumental variables* (LIV) that combines elements of conventional IV and propensity score methods (109-112). LIV methods allow for the adjustment for confounding that conventional IV offers, but also produce consistent estimates of the ATE, TT, and TUT, even under conditions of treatment

---

experimental settings because the act of randomization imposes a persistent, artificial effect on the distribution of treatments patients receive.

[7] Heckman (2007) argues that TT and TUT estimates can be constructed in randomized trials by limiting trial participation to the population subset that either currently uses or does not use a treatment, respectively. While this approach may prove practicable for social programs in which there are relatively straightforward eligibility criteria (e.g., income thresholds for Head Start), for clinical questions in which agents have discretion on treatment choice, identification *a priori* of the traits that characterize treatment utilization is much more complicated. For this reason, I would argue that even in those situations in which a randomized trial can be performed, only observational studies can measure a TT and TUT that reflects "real world" treatment choices.

effect heterogeneity. Their implementation is relatively complex, but in general, treatment effects are estimated including both observed heterogeneity (interactions between treatment choice and covariates) and unobserved heterogeneity (using a polynomial form of a propensity score, which acts as a proxy of unobserved heterogeneity). These different treatment effect values (called *marginal treatment effects*) are weighted based on the distribution of patients to produce estimates of the LATE, ATE, TT, and TUT parameters. Detailed expositions of the LIV technique are now available (113-117), but adoption of LIV has been slow in the clinical and policy communities. One reason for this delay may be the difficult nature of currently available literature on LIV, suggesting that an accessible introduction to the method's value and implementation will provide a useful contribution to the literature.

CLINICAL CONTEXT

Percutaneous coronary intervention (PCI) is an invasive procedure in which occluded coronary arteries are reopened through the use of catheters threaded into the heart. It is both commonly performed (over 1.3 million PCI procedures performed annually in the United States) and expensive (typical hospital charge of $45,000 per procedure) (118). PCI was initially performed by dilating areas of obstruction with a balloon, a procedure retroactively labeled "plain old balloon angioplasty" (119). However, damage to the vessel wall caused by the angioplasty process led to abrupt vessel closure or artery dissection in a small percentage of cases; these complications were medical crises often requiring emergent coronary artery bypass grafting. More frequently, a gradual reocclusion of the treated lesion (restenosis) led to the return of anginal symptoms requiring additional treatment. The development of the

first coronary stents (metal struts implanted at the time of angioplasty) provided dramatic reductions in the risk of emergencies such as vessel closure or dissection (120, 121). Stents also reduced the risk of restenosis by approximately 50 percent, lowering the percentage of patients requiring an additional procedure at the site of the initial procedure (target lesion revascularization) from 20-30 percent to 10-15 percent (122-125).

Numerous efforts to further reduce the rate of restenosis were tested, but most centered on reducing the proliferative response of the vascular endothelium (the interior lining of the treated coronary artery) to PCI-induced injury. While some degree of endothelial proliferation is integral to vessel healing, excessive proliferation occludes the vessel lumen. After many prominent failures (126, 127), manufacturers eventually developed Drug Eluting Stents (DES), which are coated with minuscule doses of anti-proliferative compounds such as sirolimus and paclitaxel. In a series of pivotal clinical trials, sirolimus- and paclitaxel-eluting stents were shown to dramatically reduce restenosis and the need for consequent target lesion revascularization when compared with bare metal stents (BMS), without affecting mortality (128-130). Despite being substantially more expensive than BMS, DES rapidly dominated the marketplace after their introduction in April 2003 (131).

Widespread use of DES continued through 2005 and most of 2006, but safety concerns regarding DES began to emerge. Since DES approval, there were scattered reports of DES-associated late stent thrombosis (blood clots occurring between 30 days and 1 year after insertion) (132, 133). These concerns erupted into the spotlight in September 2006 when several studies presented at the annual meeting of the European Society of Cardiology (ESC) reported an elevated risk of death in DES recipients (134-136). This increase in mortality had not been observed in clinical trials of DES and was thought to be attributable to

elevated late stent thrombosis caused by reduced adherence to antiplatelet therapy [e.g., clopidogrel (Plavix®)] (137). The Food and Drug Administration immediately launched an inquiry into DES safety. Amid widespread clinical and lay media coverage, the DES market share in the United States fell from over 90% to barely 60% by mid-2007 (138). Ultimately, it was determined that DES were safe for most patients provided they were able to remain on antiplatelet therapy for at least 12 months after PCI; DES rates began to climb again in early 2008 **(Figure 5.3)** (139).

In the immediate wake of the safety concerns regarding DES, American and Canadian investigators undertook a series of observational CER studies comparing DES and BMS. Interestingly, rather than the increased mortality seen in the studies presented in 2006 at ESC, most studies using conventional model-based or propensity score-based adjustment found strong *reductions* in mortality associated with DES use, with the relative risk reduction in all-cause mortality frequently estimated to be 25 percent or more (69, 140-143). It is possible that DES improves survival among a subset of patients who are receiving DES who receive DES in clinical practice, but who were excluded from randomized clinical trials; however, the mechanism through which survival benefits of such magnitude would be achieved is unclear. These observational analyses of DES versus BMS may also be limited by endogenous treatment selection, as DES are used preferentially for patients with less acute coronary disease (144).

Supporting this explanation, comparisons of patient outcomes before and after the introduction of DES (considered in the clinical outcomes literature to be a "semi-IV" form of analysis) failed to show a mortality benefit to DES usage (145, 146). However, these studies did not calculate the LATE, instead reporting only the effect of widespread DES usage on the

entire population of PCI recipients. These results are thus only indirect evidence of DES effectiveness, and because these studies were performed on different cohorts using different definitions than the conventional CER studies, direct comparison of results is problematic. Others have explicitly compared propensity score and instrumental variables approaches, using the same cohort of patients for both methodologies (147, 148); these studies demonstrated that even among the same cohorts of patients, conventional and IV-based approached produced markedly different estimates of DES efficacy.

However, comparing LATE estimates to the ATE and TT values produced by conventional CER methods is clearly problematic. Even with excellent control for confounding, the LATE, ATE, TT, and TUT parameters may differ from each other due to treatment effect heterogeneity rather than due to either the conventional or IV methods being the "correct" result. Intuitively, in the presence of potential treatment effect heterogeneity, comparing conventional and IV analysis involves an implicit "apples to oranges" contrast, because the two methodologies do not claim to estimate the same quantity (149). Local IV methods provide a solution to this problem, because they allow for the comparison of like estimands.

A MODEL FOR TREATMENT SELECTION AND OUTCOMES

Counterfactual-based definitions of heterogeneous treatment effects[8]

Patients experience outcomes (Y) based on the treatment they received (s) and their characteristics (ω) according to the function $Y(s,\omega)$. Both s and ω may be either a scalar or vector, in that s may contain a bundle of treatment components (e.g., repeated doses of a medication, staged surgical procedures, combinations of therapies) and that multiple patient characteristics may comprise ω (e.g., clinical, demographic, and socioeconomic factors). For simplicity, consider without loss of generalizability the case of a binary treatment choice such that $s \in \{0,1\}$. Additionally, consider ω to consist of two components: X being those factors observed by the analyst and W being those factors unobserved. As a result, Y can also be expressed as a function of s, X, and W. As with other uses of potential outcomes models, we assume independence between X and W.

For any given patient at any given time, we can only observe those outcomes associated with one value of s:

$$Y(X,W) = DY(1,X,W) + (1-D)Y(0,X,W)$$

where $D = 1$ if the patient received $s = 1$ and $D = 0$ when the patient received $s = 0$. As a result, in most circumstances it is impossible to calculate individual-level treatment effects, which are Marshallian *ceteris paribus* differences in outcome associated with the change from $s = 0$ to $s = 1$ for the individual (150). This fundamental problem in causal inference

---

[8] This section draws heavily from the exposition of Anirban Basu in "Estimating Decision-Relevant Comparative Effects Using Instrumental Variables" Stat Biosci 2011 Sep;3(1):6-27 as well as Chapters 70 and 71 (both by Heckman and Vytlacil) in the *Handbook of Econometrics*. I review the key points in a more approachable manner, but readers interested in a detailed exposition of the local instrumental methodology are urged to consult Heckman's chapters in the *Handbook* for a conceptual understanding of the LIV approach, followed by Basu's work for an illustration of its application of clinical outcomes research.

(the *evaluation problem*) has been long known to both economists and statisticians; the above model is known as both the Quandt and Roy models in economics and the Neyman-Fisher-Cox-Rubin model in statistics (151-153).

Most applied research avoids the evaluation problem by estimating summary measures (e.g., means) of population-level treatment effects instead of individual-level treatment effects. The most common of these is the ATE, which can be expressed as the mean expected difference in outcome:

$$\begin{aligned} ATE = E[Y(1) - Y(0)] &= \iint [Y(1, X, W) - Y(0, X, W)]\, dF(W|X) dF(X) \\ &= \iint \Delta(X, W)\, dF(W|X) dF(X) \end{aligned} \tag{1}$$

Intuitively, this equation states that the ATE is an average of the expected difference between treated and untreated patients, weighted by the empirical distribution of patient characteristics (both *X* and *W*) in a given sample. $\Delta(X, W)$ is defined as $[Y(1, X, W) - Y(0, X, W)]$ for the purposes of simplifying subsequent notation.

The TT and TUT parameters are also averaged differences in expected outcome, but with weighting based on the distribution of patient characteristics (*X, W*) among those patients in the population who do or do not receive treatment, respectively for TT and TUT:

$$TT = E[Y(1) - Y(0)|D = 1] = \iint \Delta(X, W) dF(W|X, D = 1) dF(X|D = 1) \tag{2}$$

$$TUT = E[Y(1) - Y(0)|D = 0] = \iint \Delta(X, W) dF(W|X, D = 0) dF(X|D = 0) \tag{3}$$

The counterfactual framework thus provides a compact mechanism useful for defining several important estimands.

With observational data, however, translating these definitions into empirical estimates is challenging. Simply calculating differences between mean values of *Y(1)* and *Y(0)* in the observed cohort will only estimate ATE, TT, and TUT if *Y(1) and Y(0)* are orthogonal to *D* ($Y(1), Y(0) \perp D$), which implies that *D* is not a function of *X* or *W*. In observational data from clinical practice, rarely are potential outcomes and treatment choices independent because agents[9] actively attempt to match patients with their optimal treatment. This issue is dubbed the *selection problem* (i.e., endogenous treatment choice or confounding by indication).

There are two primary mechanisms by which the selection problem is normally addressed in CER using observational data:

1. **Outcome or treatment modeling:** The basis of regression adjustment of either outcomes or treatment choice is that by conditioning on an observed subset of ω that is correlated with both outcome and treatment choice (X_C*;* the confounding variables), it is possible to produce independence between *Y* and *D*:

$$Y(1), Y(0) \perp D \mid X_C$$

If this relationship holds, then the remaining variation in *D* after conditioning on X_C can be considered equivalent to randomization and used for the consistent estimation of treatment effects (116). In other words, any remaining unmeasured factors (*W)* that affect outcome must be independent of *D* after conditioning by X_C. This assumption cannot be formally tested – it must instead be justified based on subject area knowledge of confounding factors and the quality of the available data.

[9] "Agents" in this setting can be patients, providers, managed care entities, insurers, or some combination of decision makers.

2. Instrumental Variables: In contrast to conditioning methods, which uncover natural randomization "by subtraction" through the identification and removal of confounding variables, IV methods rely on identifying factors associated with natural randomization (*Z*). These factors (instruments) must, after conditioning on X_{IV}, be independent with respect to outcomes (*validity criterion*) but must be associated with treatment choice (*strength criterion*):

$$Y(1), Y(0) \perp Z \mid X_{IV} \text{ and } \Pr(D = 1|Z, X_{IV}) \neq \Pr(D = 1|X_{IV})$$

The major benefit of the IV approach for some empirical questions is that it is possible to identify instruments for which X_{IV} is present in the data available to the analyst, even when X_C is not available in the data. In the extreme case of "perfect" IVs, X_{IV} can even be an empty set (*i.e.*, no conditioning on observables is necessary); the randomization in a clinical trial can be viewed as one form of perfect IV. As with the independence of *Y* and *D* in the case of conditioning on observables, the validity criterion is also untestable and requires appeals to theory and contextual knowledge in order to justify.[10]

The major limitation of IV methods is that their natural estimand is the LATE, which is a weighted average of treatment effects among only those patients who would receive the

[10] When multiple instrumental variables are identified for a given empirical question, many econometricians advocate tests of "over-identification" as a mechanism by which instrumental variable validity may be assessed. However, such tests cannot differentiate between invalid instrumental variables and a number of other circumstances, including the presence of heterogeneous treatment effects (i.e., they are sensitive for IV validity but not specific). For this reason, they are of little value in CER for clinical medicine and I will not apply them in this study.

treatment *s =0* if the IV $Z = z$ and treatment *s =1* the when $Z = z'$ (the marginal population): [11]

$$LATE = E[Y(1) - Y(0)|D(z) - D(z') = 1] = \iint \Delta(X,W)dF\big(W|X, D(z) - D(z')\big)dF\big(X|D(z) - D(z')\big) \quad (4)$$

Thus, the peculiarity of conventional IV methods is that they incorporate the tool used to identify the model (the instrument) into the definition of what the model is estimating (the LATE estimand). This conflation of model identification and estimand definition is the crux of the contentious debate between groups headed by Joshua Angrist and Guido Imbens (in favor of LATE) and James Heckman and Angus Deaton (against LATE) regarding the use of IV for treatment comparisons (154-159).[12]

Treatment effect heterogeneity affects estimation of treatment effects

Equations 1-4 demonstrate that all commonly-estimated average treatment effects can be viewed as weighted averages of *X*- and *W*-specific treatment effects ($\Delta(X,W)$, or $Y(1,X,W) - Y(0,X,W)$), which are weighted based on distributions of *X* and *W* in the population of interest. For the case of the ATE, it is the distribution of *X* and *W* among all patients in the cohort; for TT and TUT, it is the distribution among treated and untreated patients, respectively, and for LATE, it is the distribution among "marginal patients". From these definitions of the treatment effects, it is clear that ATE, TT, TUT, and LATE will be

[11] For instrumental variables with more than two values, the LATE is an information-weighted average of a series of comparisons of *z* and *z'*.

[12] The debate between the Heckman/Deaton and Angrist/Imbens factions regarding the use of instrumental variables methods even spilled into the non-technical literature; for example, see http://www.economist.com/node/14210799.

equal to each other in two circumstances. The first is when there is no heterogeneity in treatment effect (when $\Delta(X, W) = \Delta$ for all *X, W*). Intuitively, if all patients have the same response to a particular treatment (at least relative to the comparator treatment), then using different weights to average the treatment responses will have no effect on the resulting estimates. The second scenario is when there is no correlation between the treatment effect and the distributions of *X* and *W*, meaning that while $\Delta(X, W)$ may vary in the population, there is no difference in the distributions of *X* and *W* in the treated, untreated, and "marginal" populations.

The principles necessary to address heterogeneous treatment responses are well developed when heterogeneity is governed by *X,* but not *W*. In this case, $\Delta(X) = \Delta(X, W)$ for all *W*. Thus, equations 1-4 can be simplified to integrations of $\Delta(X)$ over the distribution of *X* corresponding to the estimand. Because *X* is observed in the data, such integration is straightforward to implement empirically and can be performed with matching or weighting, as is often performed for propensity score analyses (85, 160). Different weighting functions exist when the objective is to recover the ATE, TT or TUT parameters. It is important to note that the need to do some form of active adjustment to recover relevant treatment parameters is common to many regression techniques, not just IV models, when treatment effect heterogeneity is a function of *X,* and *X* is associated with treatment choice. Conventional regression methods such as ordinary least squares estimate ATE when treatment effects are constant. However, when treatment effect heterogeneity is correlated with treatment choice, even when confounding is fully-adjusted, these models instead estimate a conditional variance-weighted averaged treatment effect parameter (161).

Unfortunately, the approaches that allow for the consistent estimation of treatment effects when treatment effect heterogeneity is due to observed factors (X) fail when treatment effect heterogeneity is due to unobserved factors (W) and treatment choice is not independent of W. This limitation occurs because it is not possible to weight $\Delta(X, W)$ by the empirical distribution of W because the distribution of W is, by definition, unknown. Such a scenario is dubbed *essential heterogeneity* in the econometrics literature (109). It is likely to be common in observational CER, as physicians and patients almost always have more information than the analyst for predicting a patient's response to therapy, which influences their choice of treatment. Such concerns are particularly salient when the covariates available for performing observational CER are obtained through use of administrative or general purpose registry data rather than prospective data collection. However, by relying on choice theory regarding agent behavior, we can recover treatment effects while incorporating essential heterogeneity.

An index model of treatment choice informs the essential heterogeneity problem

Based on a generalized Roy model, Heckman and his colleagues (1999) have developed a model of treatment effects that both clarifies how the ATE, TT, and TUT parameters can be estimated from observational data and can be used to unify the methodologies underlying treatment comparisons (e.g., trials, propensity score designs, IV) (116). Here, I concentrate on explaining the model in terms of how it can be used for IV, again considering the case of a binary treatment choice without loss of generalizability.

Agents (patients, physicians, or a combination) select treatments using an index model in which observed (confounding factors X as well as instruments Z) and unobserved

(*W*) factors are used to determine whether to use the treatment based on the value of latent anticipated utility *D**:

$$D^* = \mu_D(X, Z, W) + U_{D^{*\prime}}$$

where $U_{D^{*\prime}}$ is the random error term. Under standard exogeneity assumptions treating *Z* and *X* as independent of both *W* and $U_{D^{*\prime}}$, we can define $D^* = \mu_D(X, Z) + U_{D^*}$, where $E(U_{D^*}) = 0$. Agents choose treatment 1 if $D^* > 0$ and treatment 0 if $D^* < 0$:

$$D = \mathbf{1}(D^* \geq 0) = \mathbf{1}\big(U_{D^*} > -\mu_D(X, Z)\big) \Leftrightarrow D = \mathbf{1}\left(F_{U_{D^*}}(U_{D^*}) > F_{U_{D^*}}\big(-\mu_D(X, Z)\big)\right)$$

We define *P(X,Z)* to be a propensity score predicting receipt of treatment as a function of both observable covariates and instruments such that $P(X, Z) = Pr(D = 1|X, Z) = F_{U_{D^*}}(\mu_D(X, Z))$. We define another variable, U_D, as a uniform random variable with range between zero and one ($U_D \sim$ Unif[0,1]) such that $U_D = F_{U_{D^*}}(U_{D^*})$, meaning that U_D is the probability transformation of U_{D^*}. Intuitively, *P(X, Z)* represents variation in treatment choice due to factors observed by the analyst, while U_D is variation due to unobserved factors. *P(Z)* and U_D are assumed to be independent.

Using these definitions, we can rewrite *D* accordingly:

$$D = \mathbf{1}(U_D > 1 - P(Z, X))$$

The benefit of this form is that it combines with the marginal treatment effect (introduced next) to provide treatment effect estimates.

Marginal treatment effects enable estimation of ATE, TT, and TUT

We can re-define ATE, TT, and TUT parameters in terms of the index model of treatment choice by first calculating the Marginal Treatment Effect (MTE). MTE, developed by

Björklund and Moffitt,(162) is an average treatment effect conditional on specified values of X and U_D:

$$MTE = E[\Delta|X, U_D]$$

Intuitively, this value is equal to the treatment effect that would be expected for patients with characteristics X who would be indifferent between receiving treatment and not receiving treatment if they were randomly assigned a value of instrument Z such that $U_{D^*} = \mu_D(X, Z) = 1 - P(Z, X)$. With this linkage between the unobservable (U_{D^*}) and the observable ($P(Z, X)$), the method of local instrumental variables can be used to estimate MTE from empirical data. In LIV, we estimate outcome Y as a non-linear function of X, $P(X, Z)$, and interactions between X and $P(X, Z)$. We then calculate the partial derivative of Y with respect to $P(X, Z)$, evaluated at different values of U_D:

$$\frac{dE[Y|X, Z]}{dP(X, Z)} = MTE(X, U_D)$$

where $U_D = P(X, Z)$. A non-linear functional form for $P(X, Z)$ in the outcome equation allows for MTE to vary as a function of U_D.

Heckman and colleagues developed weights for combining X- and U_D-specific values of the MTE into estimates of standard treatment effects (109):

Estimator	**Weight**
ATE	$\omega_{ATE} = 1$ (constant for all)
TT	$\omega_{TT}(X, U_D) = \frac{Pr(P(X, Z) > U_D \mid X, U_D)}{\iint \Pr(D = 1 \mid X, U_D) dU_D dF_X}$
TUT	$\omega_{TUT}(X, U_D) = \frac{Pr(P < U_D \mid X, U_D)}{\iint \Pr(D = 0 \mid X, U_D) dU_D dF_X}$

TT weighting thus upweights MTE estimates with high values of U_D, meaning that patients who possess unobserved characteristics giving them greater propensity to choose the treatment have more influence on the aggregate TT estimate than those patients with low values of U_D, while the opposite is true for TUT. However, in practice these aggregated weighting based approaches have proven problematic to implement empirically (personal communication with Anirban Basu). Instead, we rely on a recently developed approach based on Patient-Centered Treatment (PeT) effects (163).

The PeT effect framework allows for the estimation of individual-level treatment effects from the MTE. Because U_D is not observable in the data, calculating a X- and U_D-specific treatment effect is not possible. Instead, we rely on the assumptions above that: 1. U_D and X are orthogonal; 2. $U_D \sim$ Uniform(0,1); and 3. Patients choose treatment D when $U_D > 1 - P(Z,X)$. With these assumptions in place, it follows trivially that X-specific treatment effects can be obtained by simply integrating the MTE with respect to the distribution of U_D:

$$PeT(X) = \int_0^1 [MTE(X,U_D)][(D)(P(X,Z) > U_D) + (1-D)(P(X,Z) < U_D)]dU_D$$

PeT effects (which are conditional on X) can be aggregated into unconditional ATE, TT, and TUT effects by averaging over all patients, treated patients, and untreated patients, respectively:

Estimator	**Weight**
ATE	$\int PeT(X)dX$
TT	$\int PeT(X)d(X\|D=1)$

TUT $\int PeT(X)d(X|D=0)$

EMPIRICAL APPROACH

Medicare claims data

As with Study 2, this study utilizes data from the Federal Medicare program. Details regarding initial stages of dataset creation are identical to those which were reported in Study 2 (the sample size differs slightly due to a different version of the analytic dataset being used for analysis). In brief, we used the 2003-2004 Medicare Provider Analysis and Review (MedPAR) file to identify patients who were admitted for an acute coronary syndrome (either myocardial infarction or unstable angina). Using matching MedPAR, outpatient, and carrier (physician) claims for all patients, we identified the use of PCI or CABG within 30 days of an acute coronary syndrome admission. A series of exclusions were made to include only patients in which complete follow-up data were available (documented in Study 2). All outcomes were measured relative to the date of the first ACS-related revascularization procedure.

Cohort selection

In this study, the cohort was limited to individuals receiving PCI with stent insertion, rather than CABG, for their revascularization procedure. To ensure a clean comparison of DES versus BMS, we excluding patients that received both types of stents (in previous analyses, we have demonstrated our findings to be robust to defining DES based on receipt of any DES versus receipt of only DES). We also excluded a small number of patients with invalid date information, and a more substantial number of patients with missing covariate

values. This was deemed acceptable for this analysis because the primary intention is to demonstrate the use of the LIV methodology rather than to produce actionable estimates; however, in our previous studies of DES and BMS use of more complex approaches to addressing missing data (such as imputation) produced similar results.

Outcomes

We evaluated four outcomes, using definitions previously employed in the peer-reviewed literature:

1. **All-cause mortality,** measured in the Medicare denominator file.

2. **Readmission for MI,** measured from Inpatient claims in which an International Classification of Diseases, Ninth Edition, Clinical Modification (ICD-9-CM) diagnosis code for acute MI (410.x1) was included a discharge diagnosis..

3. **Readmission for major bleeding,** measured from inpatient claims in which ICD-9-CM diagnosis codes 430-432 (intracerebral hemorrhage), 478.X (gastrointestinal), 719.1X (hemarthrosis), 423.0 (hemopericardium), 599.7 (hematuria), 626.2, 626.6, 626.8, 627.0, 627.1 (vaginal), 486.3 (hemoptysis), 784.7 (epistaxis), and 459.0 (hemorrhage not otherwise specified) were included as a final diagnosis (164). Because it was not possible to determine the onset time of bleeding within an admission, only bleeding-related admissions subsequent to the index procedure discharge date were included.

4. **Inpatient repeat revascularization,** measured from inpatient and carrier claims claims using the criteria described in Study 2. Only revascularizations subsequent to the index procedure discharge date were included.

All outcomes were measured for 46 months following revascularization, the longest follow-up period available for all included patients. This restriction is suboptimal in the sense that it means that longer-term follow-up data available for patients treated early in the study period were ignored. However, the restriction was necessary because, as previously mentioned, LIV methods have not been developed to address censored data (113).

Covariates and model selection

We selected covariates thought to be predictive of differences in outcomes or in differences in response to DES. This *a priori*, theory-driven approach to covariate selection is consistent with the evolving consensus that model selection based on predictive qualities of covariates through *c*-indices or other metrics is neither theoretically sound nor empirically useful when models are constructed for the purposes of inference rather than prediction (165). Covariates included patient demographics (age, gender, race), state Medicaid buy-in status, Rural-urban commuting area, and ZIP code mean household income, obtained from the 2000 United States Census. As in Study 2, clinical covariates included whether the patient received a multiple vessel PCI procedure, the specific ACS diagnosis, and clinical comorbidities using the Elixhauser criteria. Facility medical school affiliation and ownership status were obtained from the Medicare provider of services file.

Statistical analysis

Statistical analyses were performed in SAS System, version 9.2 (SAS Institute, Cary, North Carolina), Stata/SE, version 12.1 (Statacorp, College Station, TX), and R, version 2.5.1 (R Foundation, Vienna, Austria). Programs to implement the PeT approach were

provided by Anirban Basu (personal communication with author) and were implemented in R by the author **(Appendix B)**. An alpha level of 0.05 was considered statistically significant, and two sided-statistics were employed for symmetric distributions. Institutional Review Board approval was the Institutional Review Board of the University of North Carolina at Chapel Hill.

Descriptive statistics

DES recipients were compared with BMS recipients, and patients treated in the low-DES usage facilities were compared to patients treated in high usage facilities using descriptive statistics such as median, 25th and 75th percentiles for continuous variables and frequency and percentage for categorical variables. Bivariate tests of association were conducted using Pearson chi-square tests for categorical variables and Kruskal-Wallis tests for continuous or ordinal variables.

Local Instrumental Variable Analysis

The LIV process consists of four analytic stages.

1. We first estimated the propensity score *P(X,Z)* predicting the probability that patient received drug-eluting stents.

$$P = \mathrm{f}(\alpha_0 + X\alpha_1 + Z\alpha_2)$$

After model development considering logit, probit, complementary log-log, and linear link functions, a logit link was found to offer the best fit. Unlike traditional propensity score analysis in which instrumental variables should be excluded to maximize statistical power (84), in the LIV method, IVs are included in the propensity score model. This is analogous to

the first stage equation in conventional IV formulations such as two stage least squares. We tested multiple formulations of the instrument and an indicator variable functional form, choosing the best trade-off between flexibility and parsimony of functional form using Bayesian Information Criteria.

2. Next, for each outcome, we developed models *Y(X, P(X,Z))* including *X, P(X,Z),* and interaction terms between each element of *X* and *P(X,Z).*

$$Y = \beta_0 + X\beta_1 + XP\beta_2 + P\beta_3$$

As above, we tested a variety of link functions and functional forms of covariates in an attempt to optimize model specification. We also formally tested whether essential heterogeneity existed (at link scale) and optimized the specification of the outcome equation by testing specifications of the model including higher-order (second, third, and fourth degree) polynomial terms of the propensity score in the outcome models (113). Statistical significance of the higher-order polynomial terms would indicate that essential heterogeneity is present. The polynomial formulation of highest degree that is incrementally statistically significant was used as the final specification of the model.

3. Using the specification of the outcome equations identified in step 2 (which varies for different outcomes), we next calculated the marginal treatment effect (MTE). This was performed by calculating, for each patient, estimates of MTE specific for their covariate profile (X) and for a range of potential values of U_D. Basu's program performs this step by sampling randomly (1000 draws per patient) using a uniform distribution across the range of the minimum propensity score observed in the data to the maximum propensity score. I instead present results that included 100 MTE values per patient, using a uniform range of values of UD from the lowest to the highest propensity score values in the region of common

support. I made this change both to accelerate speed of program execution and because the sampling methodology employed by Basu's program results in patients near the highest and lowest propensity scores having a high probability of not having a PeT effect estimated. I also used Basu's methods and obtained very similar results. MTE as a function of U_D is calculated by replacing P in the above equation with a range of different values for U_D and calculating the total marginal effect accordingly.

$$MTE(X, U_D) = \frac{dY(X, P)}{dP}$$

4. Finally, PeT effects were calculated for each individual patient by averaging MTE values in which U_D was greater than the observed propensity score (for DES recipients) and less than the observed propensity score (for BMS recipients). Overall average treatment effects were measured for the population by averaging the PeT effects across all (ATE), treated (TT) and untreated (TUT) patients. Because the variance theory of this estimator has not been developed, standard errors were calculated using bias-corrected bootstrapping methods using 500 iterations (166, 167). These results were compared with more traditional IV results estimated by standard bivariate probit approaches as well as a standard adjusted model using probit regression.

RESULTS

Study population

In total, 69,740 patients in the Medicare dataset were treated from May 1, 2003 and February 28, 2004 using a PCI with stent insertion procedure **(Figure 5.2)**. Exclusions removed patients who received both DES and BMS (N = 3,019), had an invalid age or death

date (N = 129), or had a missing value for a key covariate (N = 6,856). Ultimately, 59,466 patients were included in the analysis.

Temporal variation in DES usage, and c*orrelations between observed covariates and both DES usage and month of PCI.*

Among the 10 months included in the analysis, the percentage of patients treated with DES varied from 26.1% to 55.6% **(Figure 5.3)**. Statistically significant differences were observed in patient, procedural, and facility characteristics based both on whether the patient received DES or BMS and whether the patient was treated in the first five months or second five months of the study period **(Table 5.1).** However, in general, time of PCI was associated with reduced imbalance of covariates **(Figure 5.5)**. Furthermore, models which compared the expected event rates in DES and BMS recipients, and patients treated in the first and second five month periods usage facilities, demonstrate that inter-temporal variation in patient risk is more modest than the variation between BMS and DES recipients **(Figure 5.6)**. For this reason, we would argue that the IV strategy employed here is likely an improvement over conventional adjustment methodologies.

Conventional modeling and IV

Using conventional multivariate adjustment with a probit model, DES usage was associated with a significant reduction in all measured outcomes, except bleeding-related readmission **(Table 5.2)**. In contrast, using a bivariate probit model (a form of IV) only reductions in repeat revascularization and MI-related readmission were significant. Most notably, the difference in mortality associated with DES usage was not statistically significant in the bivariate (IV) probit model, as well as quite quantitatively different than the

adjusted probit estimate (incremental increase of 0.2 percentage point (CI: -3.4, 3.1) versus incremental decrease of 3.9 percentage points (CI: -4.5, -3.2).

PeT-based modeling

Statistical testing for essential heterogeneity (at logit) scale using likelihood ratio tests did not produce evidence of heterogeneity in any measured outcomes **(Table 5.4)**. Thus, a linear specification of propensity score was included in the outcome equations for all outcomes. In the overall sample, average PeT effects were only significant for MI-related readmission and repeat revascularization. A statistically significant difference was noted in which treated patients (TT) experienced a more favorable result than would be expected from the untreated patients (TUT) for all outcomes except mortality. A technical concern emerged in that replicates from the bootstrapping process did not appear to approximate distributions centered at the point estimates for the various parameters **(Figure 5.7)**. We also examined the joint distribution of individual-level PeT effects **(Figure 5.8).** Correlations between individual outcomes were all modest (Spearman's rho ≤ 0.1), indicating that individual-specific treatment effects were not correlated across outcomes (e.g., patients who benefitted from DES the most in terms of repeat revascularization reductions were not those in whom a mortality benefit was also similarly large).

DISCUSSION

In this study, we demonstrated the application of clinically relevant causal estimands estimated from IV using PeT effects, a recently-developed econometric methodology to a question in cardiovascular outcomes research (whether to use drug-eluting or bare metal

coronary stents for PCI). Using Medicare administrative data, we found evidence that DES were generally safe and effective among the populations that were treated during the study period (2003-2004), but that the safety and efficacy profile of DES were considerably less favorable among the patients who received BMS instead.

Clinically, these results suggest that in 2003-2004, providers and patients applied a selective, deliberative approach to stent selection in which those receiving DES were those that would be expected to benefit from them. In contrast, it appears that patients who received BMS during the study period were those for whom the use of DES was predicted to result in less benefit in terms of reducing repeat revascularization, but potentially greater risk, particularly of bleeding-related complications. DES utilization rates have increased substantially since the study period, which raises concern whether the "marginal" patients (patients who would have received BMS in 2003 who receive DES in 2004) are being harmed by this change in practice pattern. Whether the current generation of DES and newer antiplatelet agents as well as changes in the profile of patients receiving PCI has changed the optimal choice of stent is also unclear. Together, these results suggest that while DES are a reasonable option for many patients and clearly inappropriate for others, further research is needed to understand the optimal treatment strategy on the margin of current clinical practice.

Technically, this report demonstrates that LIV-based approaches are feasible to implement in standard statistical software (Stata programs available from Basu on request), and an implementation of the PeT methodology developed by this author is provided in the R programming language as an appendix to this publication. However, further study is needed to understand the properties of these estimators. In particular, the difference between the

point estimates and bootstrapped replicates observed in this study require further investigation regarding whether the estimator underlying LIV has adequate asymptotic consistency for empirical work in which small differences in clinical effectiveness are important.

LIV only identifies parameters over the range of support provided by the first-stage propensity score. In other words, the propensity score model we construct must include patients in both the DES and BMS treatment groups who have very high and very low propensity scores (i.e., close to 0 and 1), and consensus thresholds for "very high" and "very low" are elusive. In this case, the range of common support observed in the propensity score (0.09 – 0.82) was considered adequate to enable identification of effect estimates. Future research could attempt to estimate support across the entire range of $P(X, Z)$ using extrapolation, or via placement of bounds on the treatment effects as previously described. Alternatively, estimates can be characterized as the "empirical" ATE, TT, and TUT parameters as has been done previously (112, 113).

Instrumental variables methods (both IV and LIV) are important tools for observational CER. While further work is necessary to better understand their optimal use, local instrumental variables-based methodologies show promise for allowing the estimation of clinically useful estimates from IV analyses.

TABLES

Table 5.1: Baseline patient characteristics

	Overall (N=59,466)	Type of Stent Received			Date of PCI Procedure		
		BMS (N=33,656)	DES (N= 25,810)	p	< Nov. 2003 (N= 29,445)	≥ Nov. 2003 (N= 30,021)	p
Outcomes							
Death	25.0	28.1	21.0	<0.001	24.6	25.4	0.03
Repeat Revascularization	27.6	28.6	26.2	<0.001	28.1	27.0	0.003
MI Readmission	7.9	8.4	7.4	<0.001	8.1	7.7	0.07
Bleeding readmission	3.9	3.9	3.8	0.24	3.9	3.9	0.95
Demographics							
Patient age (years)	75.8 (6.4)	76.0 (6.5)	75.5 (6.2)	<0.001	75.7 (6.4)	75.8 (6.4)	0.06
Non-white race	7.6	7.7	7.5	0.49	7.7	7.5	0.30
Female	49.6	48.5	51	<0.001	49.6	49.5	0.81
State Medicaid Buy-in	10.4	10.8	9.9	<0.001	10.6	10.3	0.34
ZIP Code Household income ($1000s)	43.1 (15.9)	42.2 (15.4)	44.3 (16.6)	<0.001	43.1 (16.0)	43.2 (15.9)	0.50
Rural-Urban Commuting Area				<0.001			0.96
Metropolitan	67.9	66.6	69.6		67.8	68.0	
Micropolitan	14.9	15.2	14.4		14.9	14.8	
Small Town	9.5	10	8.8		9.5	9.4	
Rural	7.8	8.2	7.3		7.8	7.8	
Clinical Characteristics							
Acute Coronary Syndrome Diagnosis				<0.001			<0.001

ST-elevation MI	35.1	42.4	25.6		34.4	35.7	
Non ST-elevation MI	26.3	25.3	27.4		25.3	27.1	
Unstable angina	38.7	32.3	47		40.2	37.1	
Multivessel PCI	18.4	17.3	19.8	<0.001	18.4	18.5	0.63
Comorbid Conditions							
Heart Failure	7.9	8.1	7.7	0.06	8.1	7.8	0.15
Valvular disease	4.4	4.2	4.7	0.008	4.4	4.5	0.76
Pulmonary circulatory disease	2.6	2.7	2.4	0.02	2.5	2.7	0.28
Peripheral vascular disease	14.1	14.1	14.2	0.56	13.9	14.3	0.19
Paralysis	1.4	1.5	1.2	<0.001	1.4	1.3	0.09
Neurological disease	2.7	2.8	2.6	0.03	2.6	2.8	0.06
Chronic lung disease	23.1	24.3	21.5	<0.001	22.6	23.6	0.003
Diabetes mellitus	28.5	27.1	30.3	<0.001	28.6	28.4	0.61
Hypothyroidism	12.7	12.2	13.2	<0.001	12.8	12.5	0.38
Renal failure	1.9	1.8	2.0	0.06	1.8	2.0	0.11
Cancer (any)	8.5	8.9	7.9	<0.001	8.5	8.5	0.90
Coagulopathy	1.1	1.1	1.0	0.46	1.0	1.1	0.34
Arthritis	3.8	3.8	3.9	0.50	3.7	4.0	0.05
Obesity	6.4	6.2	6.7	0.01	6.5	6.3	0.19
Weight loss	1.4	1.5	1.1	<0.001	1.3	1.4	0.07
Electrolyte abnormality	13.8	14.9	12.3	<0.001	13.5	14	0.07
Anemia	13.8	14.3	13.2	<0.001	13.5	14.1	0.05
Other Psychiatric Disorder	2.3	2.3	2.2	0.40	2.3	2.3	0.58
Depression	6.0	5.8	6.2	0.02	6.1	5.9	0.41
Hypertension w/ complications	75.6	74.1	77.6	<0.001	75.7	75.6	0.74
Facility and geographic characteristics							
Major medical school	31.7	28.9	35.4	<0.001	32.0	31.4	0.11

affiliation							
Facility type				<0.001			0.02
Non-Profit	79.8	78.5	81.5		80.2	79.4	
For-Profit	10.8	11.8	9.5		10.6	11.0	
Government	9.4	9.8	9.0		9.2	9.7	
Census Region				<0.001			0.28
Midwest	29.4	31.1	27.2		29.7	29.1	
South	39.8	40.8	38.6		39.8	39.8	
West	13.4	13.3	13.5		13.3	13.4	
Northeast	17.4	14.9	20.8		17.2	17.7	

Table 5.2: Average marginal effects from probit and bivariate probit (IV) models

	Incremental Difference in Percentage Points (DES – BMS)	
	Probit	Bivariate Probit
Death	-3.9***	-0.2
	(-4.5,-3.2)	(-3.4, 3.1)
Repeat Revascularization	-3.0***	-7.4***
	(-3.7, -2.2)	(-10.9, -3.8)
MI Readmission	-0.7**	-3.0**
	(-1.2,-0.3)	(-5.1, -0.9)
Bleeding Readmission	-0.1	-0.4
	(-0.5, 0.2)	(-1.9, 1.2)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.3: Complete probit and bivariate probit model results

	Death			Repeat Revascularization		MI Readmission		Bleeding Readmission	
	Probit: Outcome	Bivariate Probit: Treatment	Bivariate Probit: Outcome	Probit: Outcome	Bivariate Probit: Outcome	Probit: Outcome	Bivariate Probit: Outcome	Probit: Outcome	Bivariate Probit: Outcome
desonly	-0.143***		-0.006	-0.091***	-0.226***	-0.051**	-0.210**	-0.017	-0.044
	(0.012)		(0.062)	(0.011)	(0.056)	(0.016)	(0.075)	(0.020)	(0.096)
age_70	0.041***	-0.004	0.041***	0.000	-0.001	0.009	0.008	0.022	0.022
	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)	(0.010)	(0.010)	(0.013)	(0.013)
age_75	0.036***	0.000	0.036***	-0.012**	-0.012**	0.011	0.011	0.022**	0.022**
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.009)	(0.009)
age_80	0.037***	-0.016***	0.038***	-0.017***	-0.018***	-0.002	-0.003	0.018*	0.017*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	(0.008)
age_85	0.060***	-0.004	0.060***	-0.025***	-0.025***	0.029***	0.029***	0.017	0.017
	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)	(0.009)
age_85up	0.060***	-0.036***	0.062***	-0.036***	-0.038***	0.015	0.013	0.001	0.000
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)	(0.010)
nonwhite	0.069**	0.004	0.069**	-0.012	-0.012	0.053	0.053	-0.038	-0.037
	(0.023)	(0.021)	(0.023)	(0.022)	(0.022)	(0.029)	(0.029)	(0.039)	(0.039)
female	-0.094***	0.078***	-0.097***	-0.053***	-0.049***	-0.021	-0.017	0.000	0.001
	(0.013)	(0.011)	(0.013)	(0.012)	(0.012)	(0.016)	(0.016)	(0.020)	(0.021)
inc_100	-0.032	-0.027	-0.031	-0.002	-0.003	-0.022	-0.023	-0.030	-0.030
	(0.022)	(0.021)	(0.022)	(0.021)	(0.021)	(0.026)	(0.026)	(0.031)	(0.031)
inc_200	-0.007***	0.012***	-0.008***	-0.004	-0.003	-0.001	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)

inc_300	-0.002	0.003**	-0.002	-0.001	-0.001	-0.001	-0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
inc_400	-0.003	0.004*	-0.004*	0.004*	0.004*	-0.001	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
inc_500up	0.000	0.001	0.000	-0.001	-0.001	-0.005*	-0.005*	0.003	0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
dual_elig	0.201***	-0.050**	0.203***	0.001	-0.002	0.102***	0.098***	0.060	0.059
	(0.020)	(0.019)	(0.020)	(0.019)	(0.019)	(0.025)	(0.025)	(0.032)	(0.032)
micropolitan	-0.048**	0.010	-0.048**	-0.042*	-0.041*	0.033	0.033	0.019	0.020
	(0.018)	(0.016)	(0.018)	(0.017)	(0.017)	(0.023)	(0.023)	(0.030)	(0.030)
smalltown	-0.034	0.003	-0.033	-0.063**	-0.063**	0.003	0.003	0.108**	0.108**
	(0.022)	(0.020)	(0.022)	(0.021)	(0.021)	(0.028)	(0.028)	(0.034)	(0.034)
rural	-0.054*	0.014	-0.055*	-0.031	-0.030	0.037	0.038	0.084*	0.084*
	(0.024)	(0.022)	(0.024)	(0.022)	(0.022)	(0.030)	(0.030)	(0.038)	(0.038)
medsch	-0.003	0.156***	-0.010	-0.030*	-0.023	0.016	0.024	-0.010	-0.009
	(0.014)	(0.012)	(0.014)	(0.013)	(0.013)	(0.018)	(0.018)	(0.022)	(0.023)
forprofit	0.007	-0.131***	0.013	0.035	0.029	0.054*	0.046	-0.046	-0.047
	(0.020)	(0.018)	(0.020)	(0.018)	(0.019)	(0.025)	(0.025)	(0.033)	(0.033)
govt	0.019	-0.090***	0.023	0.037	0.033	-0.024	-0.029	-0.026	-0.027
	(0.021)	(0.019)	(0.021)	(0.020)	(0.020)	(0.028)	(0.028)	(0.035)	(0.035)
midwest	-0.022	-0.099***	-0.017	0.053***	0.048***	0.010	0.004	0.025	0.024
	(0.016)	(0.014)	(0.016)	(0.014)	(0.014)	(0.020)	(0.020)	(0.025)	(0.025)
northeast	-0.057**	0.130***	-0.064***	-0.035	-0.028	0.000	0.008	-0.006	-0.004

	(0.019)	(0.017)	(0.019)	(0.018)	(0.018)	(0.024)	(0.025)	(0.031)	(0.031)
west	-0.036	0.048**	-0.038*	-0.038*	-0.036*	-0.032	-0.029	-0.115***	-0.115***
	(0.019)	(0.017)	(0.019)	(0.018)	(0.018)	(0.025)	(0.025)	(0.033)	(0.033)
stemi	0.319***	-0.565***	0.346***	-0.003	-0.031	0.115***	0.082***	-0.079***	-0.085**
	(0.015)	(0.013)	(0.019)	(0.013)	(0.017)	(0.019)	(0.024)	(0.024)	(0.030)
nstemi	0.262***	-0.201***	0.271***	-0.045**	-0.054***	0.202***	0.190***	0.003	0.001
	(0.015)	(0.013)	(0.016)	(0.014)	(0.015)	(0.019)	(0.020)	(0.024)	(0.025)
pci_multv	0.054***	0.085***	0.050**	0.029*	0.033*	0.093***	0.098***	0.007	0.008
	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)	(0.019)	(0.019)	(0.025)	(0.025)
chf	0.442***	-0.035	0.443***	-0.119***	-0.122***	0.084**	0.081**	0.168***	0.167***
	(0.022)	(0.021)	(0.022)	(0.023)	(0.023)	(0.027)	(0.027)	(0.032)	(0.032)
valve	0.155***	0.032	0.152***	-0.010	-0.009	0.016	0.018	0.098*	0.098*
	(0.028)	(0.027)	(0.028)	(0.028)	(0.028)	(0.036)	(0.036)	(0.041)	(0.041)
	(0.035)	(0.034)	(0.035)	(0.037)	(0.037)	(0.045)	(0.045)	(0.053)	(0.054)
perivasc	0.210***	-0.023	0.211***	0.017	0.016	0.141***	0.140***	0.078**	0.078**
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.021)	(0.021)	(0.026)	(0.026)
para	0.323***	-0.096*	0.328***	-0.113*	-0.118*	0.210***	0.202***	0.111	0.110
	(0.047)	(0.047)	(0.047)	(0.050)	(0.050)	(0.057)	(0.057)	(0.073)	(0.073)
neuro	0.371***	-0.061	0.373***	-0.145***	-0.147***	-0.009	-0.012	0.022	0.021
	(0.034)	(0.033)	(0.034)	(0.036)	(0.036)	(0.046)	(0.046)	(0.055)	(0.055)
chrnlung	0.424***	-0.091***	0.427***	-0.057***	-0.061***	0.113***	0.108***	0.121***	0.120***
	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)	(0.018)	(0.018)	(0.022)	(0.022)

diabetes	0.240***	0.075***	0.236***	0.118***	0.122***	0.233***	0.236***	0.067**	0.068**
	(0.013)	(0.012)	(0.014)	(0.013)	(0.013)	(0.017)	(0.017)	(0.022)	(0.022)
hypothy	-0.095***	0.026	-0.096***	0.041*	0.042*	-0.018	-0.017	-0.014	-0.014
	(0.019)	(0.016)	(0.018)	(0.017)	(0.017)	(0.023)	(0.023)	(0.029)	(0.029)
renlfail	0.284***	0.089*	0.279***	-0.071	-0.066	0.072	0.077	0.024	0.025
	(0.041)	(0.040)	(0.041)	(0.043)	(0.043)	(0.050)	(0.050)	(0.061)	(0.061)
cancer	0.415***	-0.076***	0.417***	-0.038	-0.042*	-0.061*	-0.065*	0.090**	0.089**
	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)	(0.028)	(0.028)	(0.032)	(0.033)
coag	0.239***	-0.028	0.240***	-0.056	-0.057	-0.136	-0.137	0.041	0.041
	(0.054)	(0.052)	(0.054)	(0.056)	(0.056)	(0.075)	(0.075)	(0.081)	(0.081)
arth	0.177***	0.014	0.176***	-0.008	-0.007	0.064	0.065	0.095*	0.095*
	(0.030)	(0.028)	(0.030)	(0.029)	(0.029)	(0.038)	(0.038)	(0.046)	(0.046)
obese	-0.097***	-0.010	-0.097***	0.018	0.018	-0.006	-0.006	-0.025	-0.025
	(0.026)	(0.022)	(0.026)	(0.023)	(0.023)	(0.031)	(0.031)	(0.041)	(0.041)
wghtloss	0.453***	-0.060	0.454***	-0.075	-0.077	0.053	0.050	-0.033	-0.033
	(0.047)	(0.047)	(0.047)	(0.051)	(0.051)	(0.061)	(0.060)	(0.076)	(0.076)
lytes	0.304***	-0.078***	0.307***	-0.118***	-0.122***	0.009	0.005	0.050	0.050
	(0.017)	(0.016)	(0.017)	(0.017)	(0.017)	(0.022)	(0.022)	(0.027)	(0.027)
anemia	0.182***	-0.029	0.182***	0.011	0.010	0.088***	0.087***	0.225***	0.225***
	(0.017)	(0.016)	(0.017)	(0.017)	(0.017)	(0.022)	(0.022)	(0.025)	(0.025)
psych	0.128***	-0.021	0.129***	-0.199***	-0.200***	-0.047	-0.048	0.087	0.087
	(0.038)	(0.036)	(0.038)	(0.040)	(0.040)	(0.051)	(0.051)	(0.059)	(0.059)
depress	0.083***	0.031	0.081**	0.007	0.009	0.044	0.045	0.103**	0.103**
	(0.025)	(0.023)	(0.025)	(0.024)	(0.024)	(0.032)	(0.032)	(0.038)	(0.038)

htn_c	-0.075***	0.035**	-0.076***	0.072***	0.073***	0.058**	0.060**	0.004	0.004
	(0.015)	(0.013)	(0.015)	(0.013)	(0.013)	(0.019)	(0.019)	(0.024)	(0.024)
2.revascmonth		0.161***							
		(0.024)							
3.revascmonth		0.323***							
		(0.024)							
4.revascmonth		0.490***							
		(0.025)							
		(0.024)							
6.revascmonth		0.714***							
		(0.024)							
7.revascmonth		0.569***							
		(0.024)							
8.revascmonth		0.667***							
		(0.024)							
9.revascmonth		0.785***							
		(0.024)							
10.revascmonth		0.833***							
		(0.024)							
constant	-0.893*	(0.188)	-0.971*	(0.356)	(0.276)	-1.409**	-1.312**	-1.586**	-1.570**
	(0.379)	(0.371)	(0.380)	(0.370)	(0.371)	(0.455)	(0.457)	(0.530)	(0.533)

Note: While biprobit probit regressions are solved jointly and thus treatment regression equations are slightly different for

each outcome, in the interest of brevity, the treatment regression results are only displayed for the death outcome.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.4: Tests for essential heterogeneity (at raw scale)

	p-value				
	Squared	Cubic		Quartic	
		Incremental	Joint	Incremental	Joint
Death	0.94	0.44	0.20	0.52	0.43
Repeat Revascularization	0.17	0.32	0.50	0.45	0.56
MI Readmission	0.71	0.75	0.50	0.85	0.65
Bleeding Readmission	0.49	0.63	0.50	0.40	0.15

Table 5.5: Mean treatment effects estimated using PeT methodology

	Incremental Difference in Percentage Points (DES – BMS)			
	Average Treatment Effect (ATE)	Average Treatment Effect on the Treated (TT)	Average Treatment Effect on the Untreated (TUT)	Sorting gain (TT-TUT)
Death				
Overall	-0.2 (-3.5, 3.2)	-0.4 (-3.6, 2.9)	0.0 (-3.5, 3.4)	-0.5 (-1.2, 0.9)
Myocardial Infarction	0.8 (-3.6, 5.5)	0.9 (-3.1, 5.6)	0.7 (-3.8, 5.7)	0.2 (-0.7, 1.5)
Unstable Angina	-1.8 (-5.5, 2.4)	-2.0 (-5.8, 2.0)	-1.5 (-5.6, 2.7)	-0.5 (-0.8, 0.2)
Repeat Revascularization				
Overall	-6.8 (-10.4, -3.4)	-8.1 (-11.8, -4.2)	-5.8 (-9.8, -2.8)	-2.3 (-3.0, -0.9)
Myocardial Infarction	-2.8 (-8.2, 1.1)	-2.9 (-8.3, 0.9)	-2.7 (-8.0, 1.4)	-0.2 (-0.8, 1.3)
Unstable Angina	-13.2 (-18.6, -8.4)	-14.0 (-19.5, -8.0)	-12.4 (-17.7, -7.9)	-1.6 (-2.2, -0.5)
MI Readmission				
Overall	-2.9 (-5.3, -0.7)	-3.5 (-5.7, -0.9)	-2.5 (-4.6, -0.7)	-1.0 (-0.9, -0.3)
Myocardial Infarction	-3.5 (-6.4, -0.9)	-4.4 (-7.7, -1.4)	-3.0 (-5.7, -1.0)	-1.4 (-1.5, -0.8)
Unstable Angina	-1.9 (-4.5, 1.2)	-2.5 (-5.3, 0.8)	-1.3 (-3.6, 1.3)	-1.2 (-1.1, -0.9)
Bleeding Readmission				
Overall	0.4 (-1.5, 2.1)	-1.1 (-2.7, 0.9)	1.5 (-0.4, 2.9)	-2.7 (-2.7, -2.2)
Myocardial Infarction	0.5 (-2.5, 2.5)	-1.3 (-3.1, 1.4)	1.5 (-1.5, 3.1)	-2.8 (-2.7, -1.9)
Unstable Angina	0.3 (-2.2, 2.8)	-1.0 (-2.9, 1.3)	1.6 (-0.8, 4.0)	-2.6 (-2.7, -1.9)

FIGURES

Figure 5.1: Heterogeneity diagram

Figure displays interpretation of treatment effects under different combinations of heterogeneity and treatment selection.

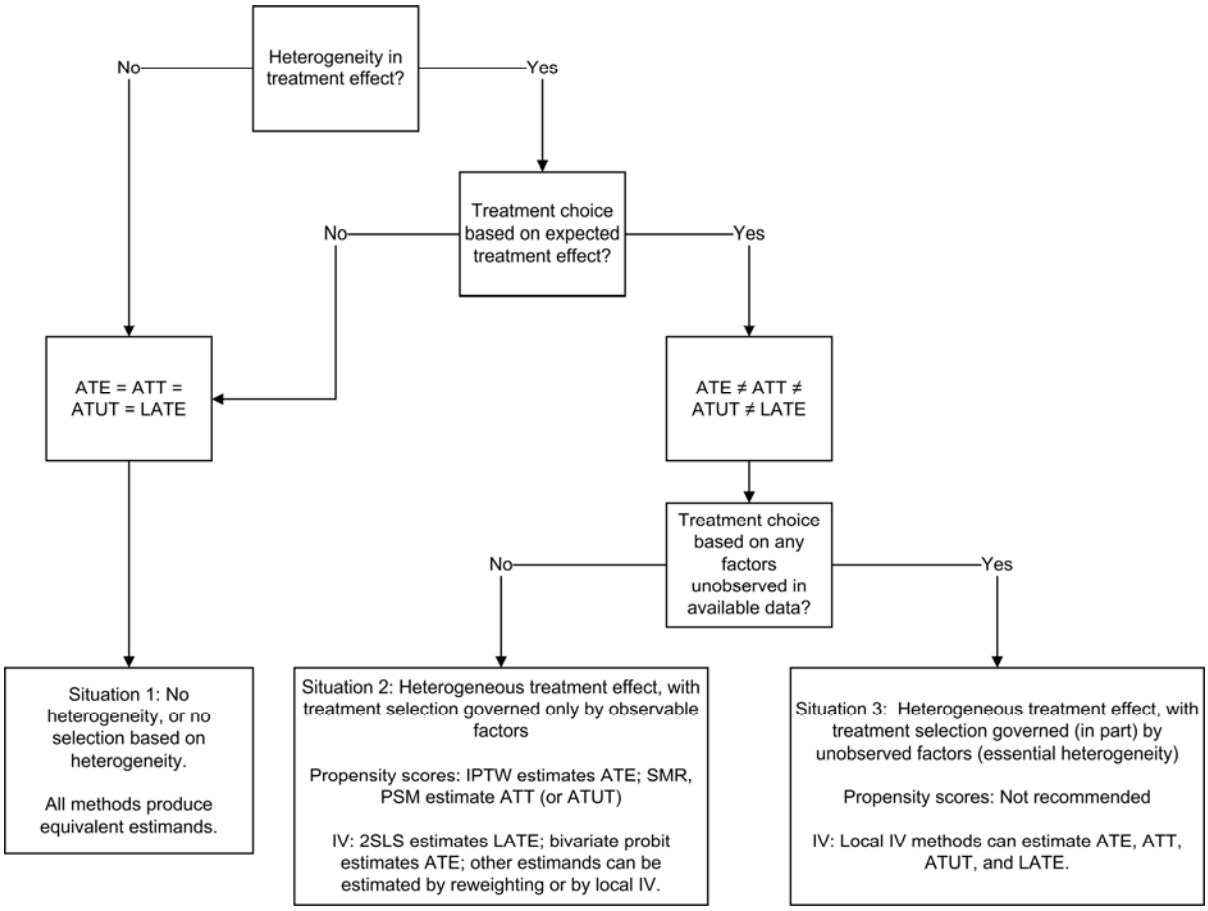


Figure 5.2: Cohort flow diagram

Figure displays derivation of final study population from the Medicare cohort.

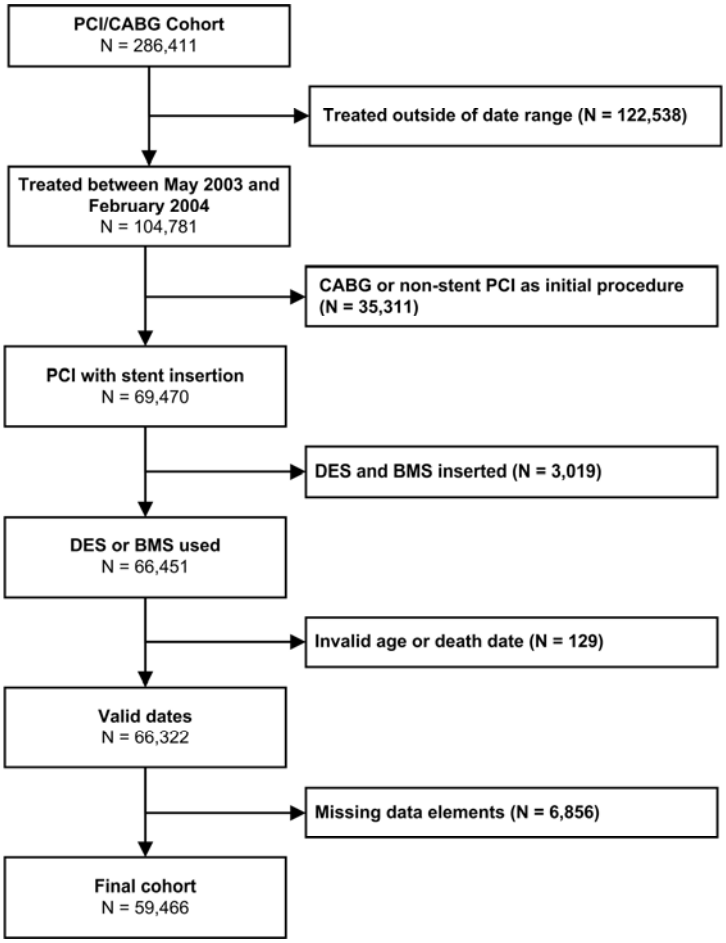


Figure 5.3: Temporal trend in DES utilization in cohort

Figure displays increasing use of drug-eluting stents over the study period.

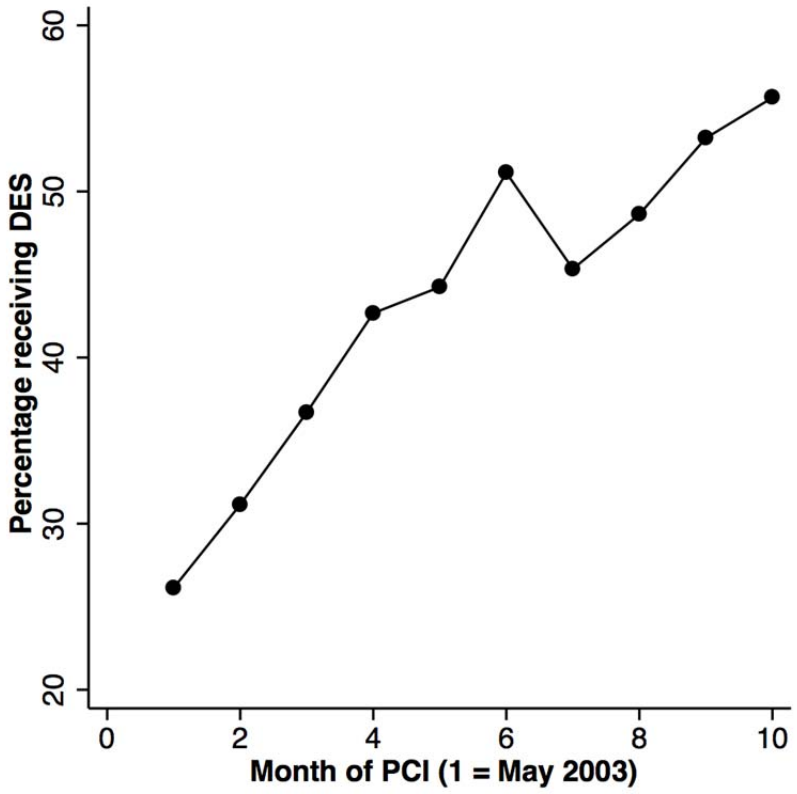


Figure 5.4: Unadjusted cumulative incidence of outcomes

Cumulative incidence plots display outcomes up to 46 months after revascularization, stratified by whether patient received drug-eluting stents (DES) or bare metal stents (BMS)

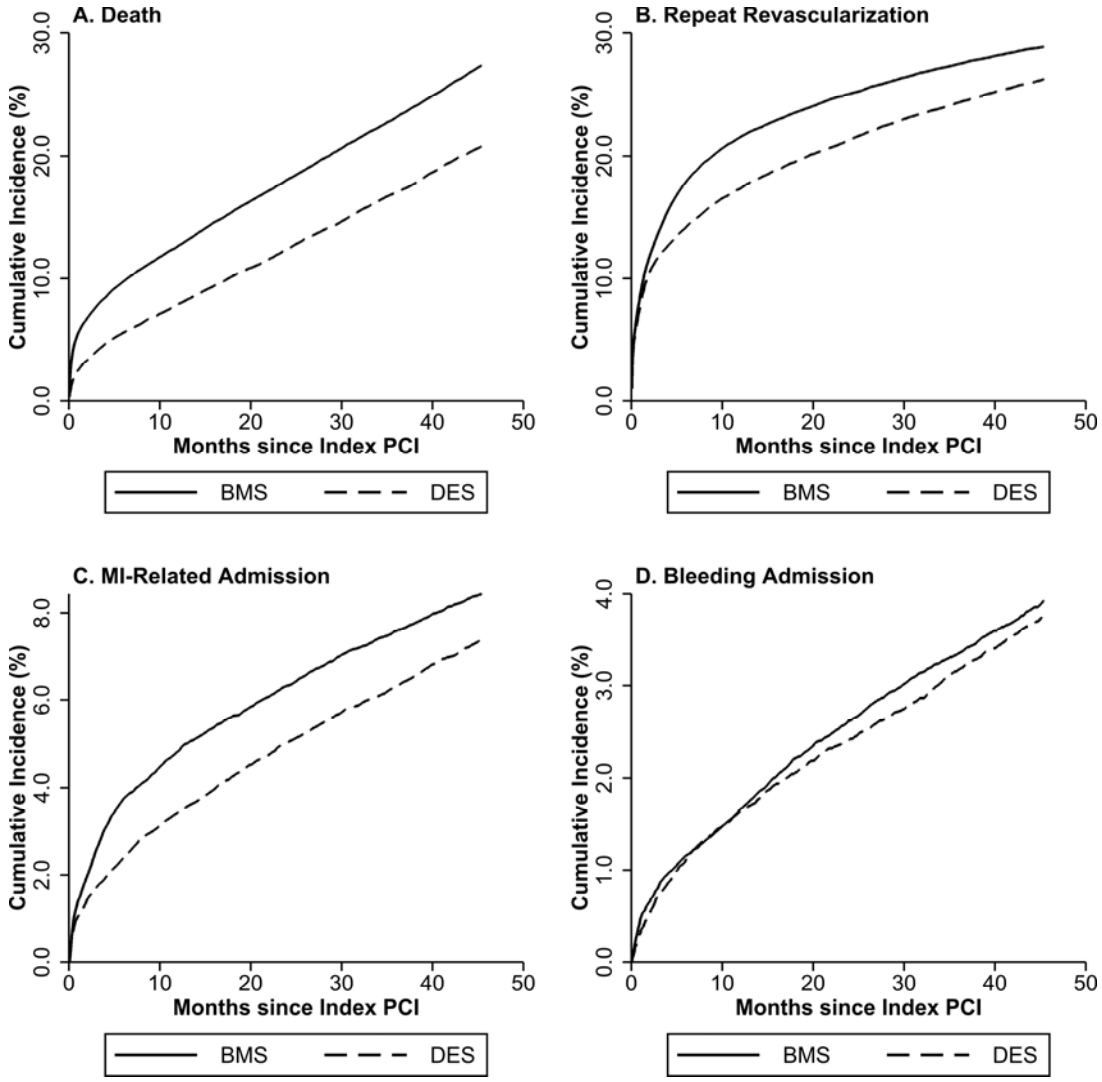


Figure 5.5: Standardized differences in baseline patient, procedural, and facility characteristics

Figure displays absolute standardized difference for covariates thought to potentially confound the relationship between stent use and outcome. Standardized differences were calculated comparing drug eluting and bare metal stent recipients, and patients treated in the first and second five months of drug eluting stent utilization.

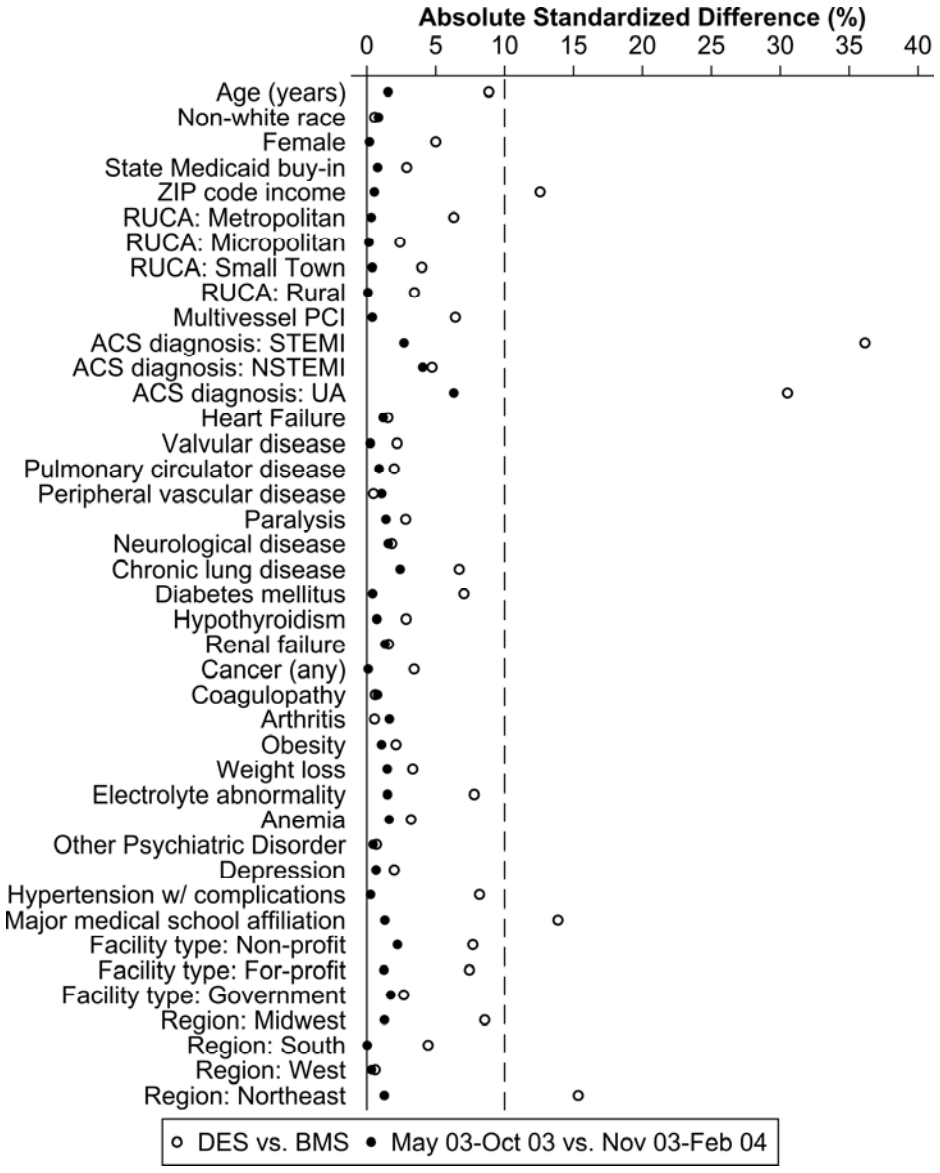


Figure 5.6: Predicted event rate based on observed baseline covariates, stratified by DES receipt and time of PCI.

Logistic regressions were fit, predicting probability of each event as a function of observed covariates but excluding DES versus BMS utilization and time of PCI. Logistic models were then used to calculate expected event rate in DES vs. BMS recipients, and first and second five month periods.

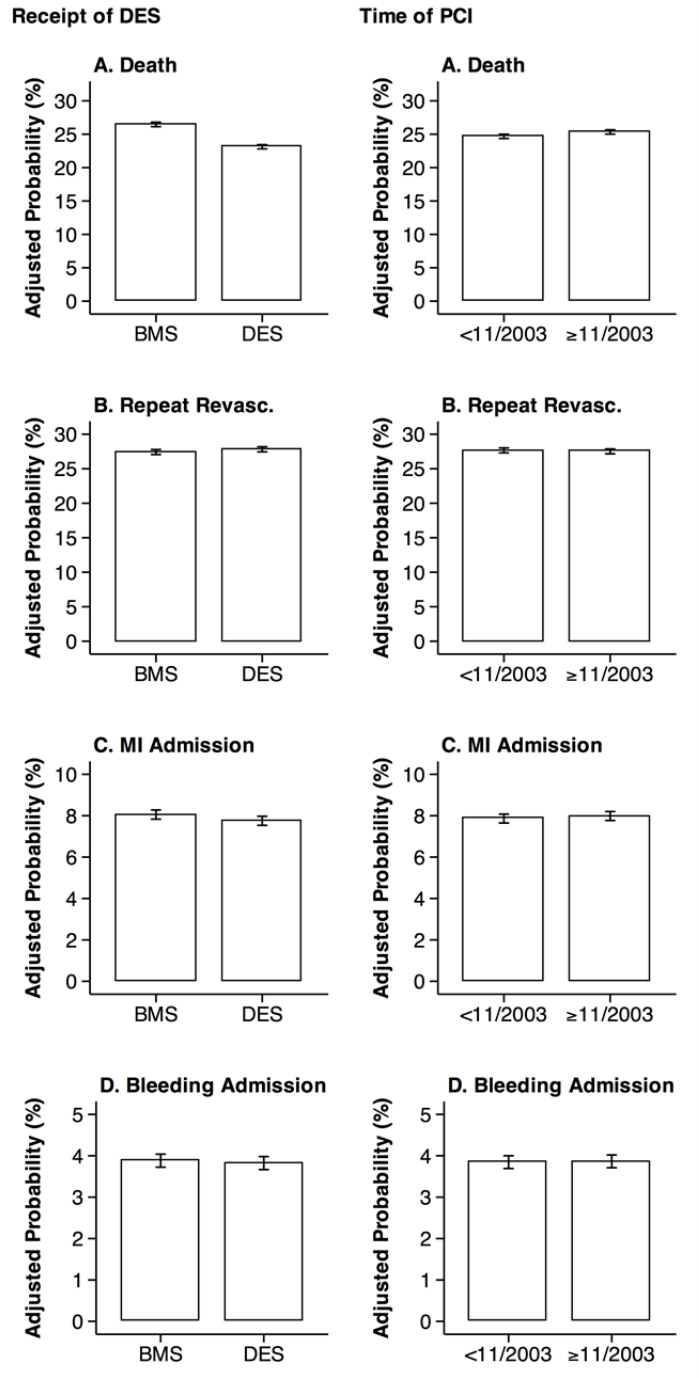


Figure 5.7: Comparison of bootstrapped replicates to point estimates.

Histograms display results from bootstrap iterations. Dashed lines indicate the point estimate from the original sample.

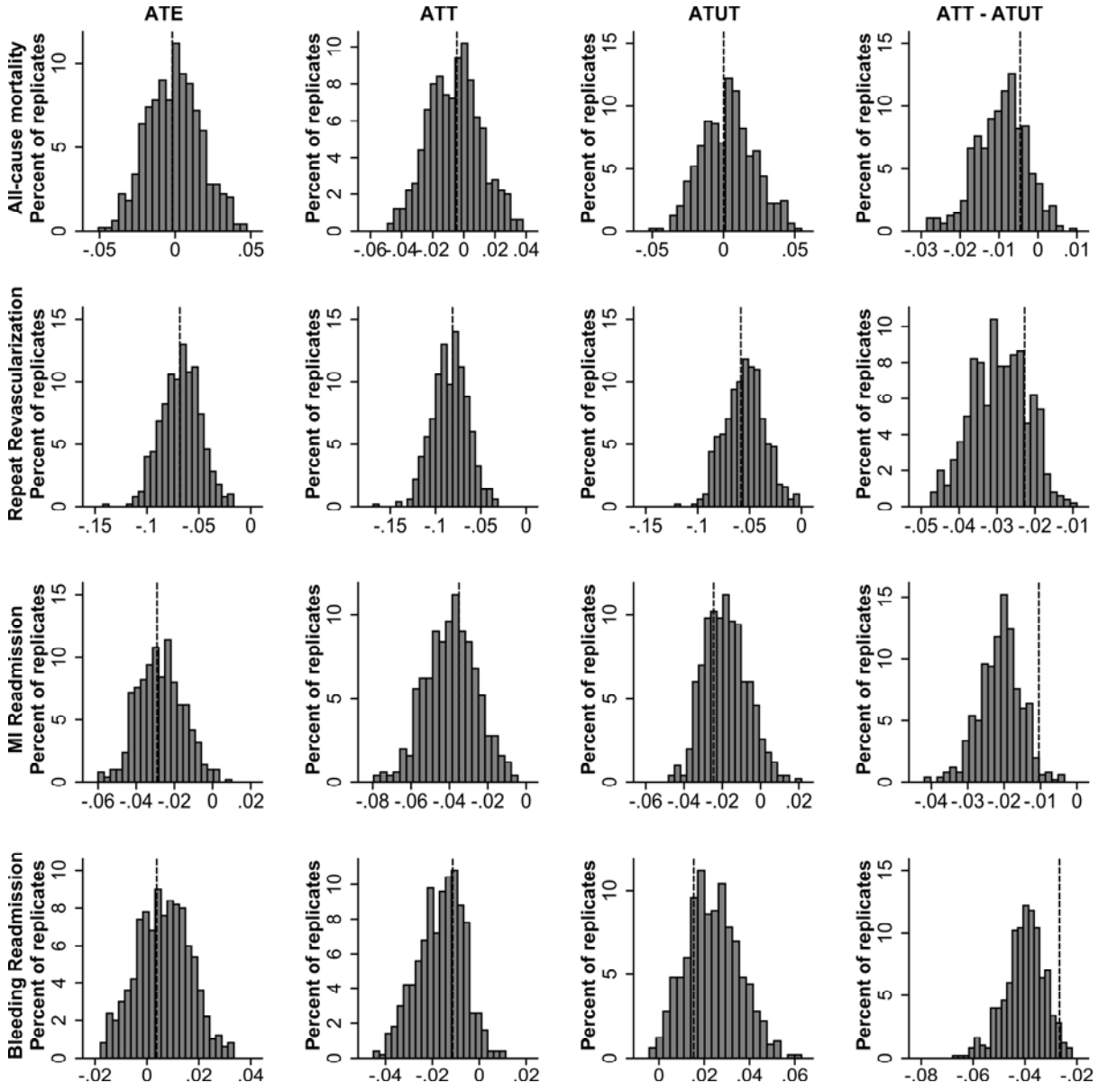
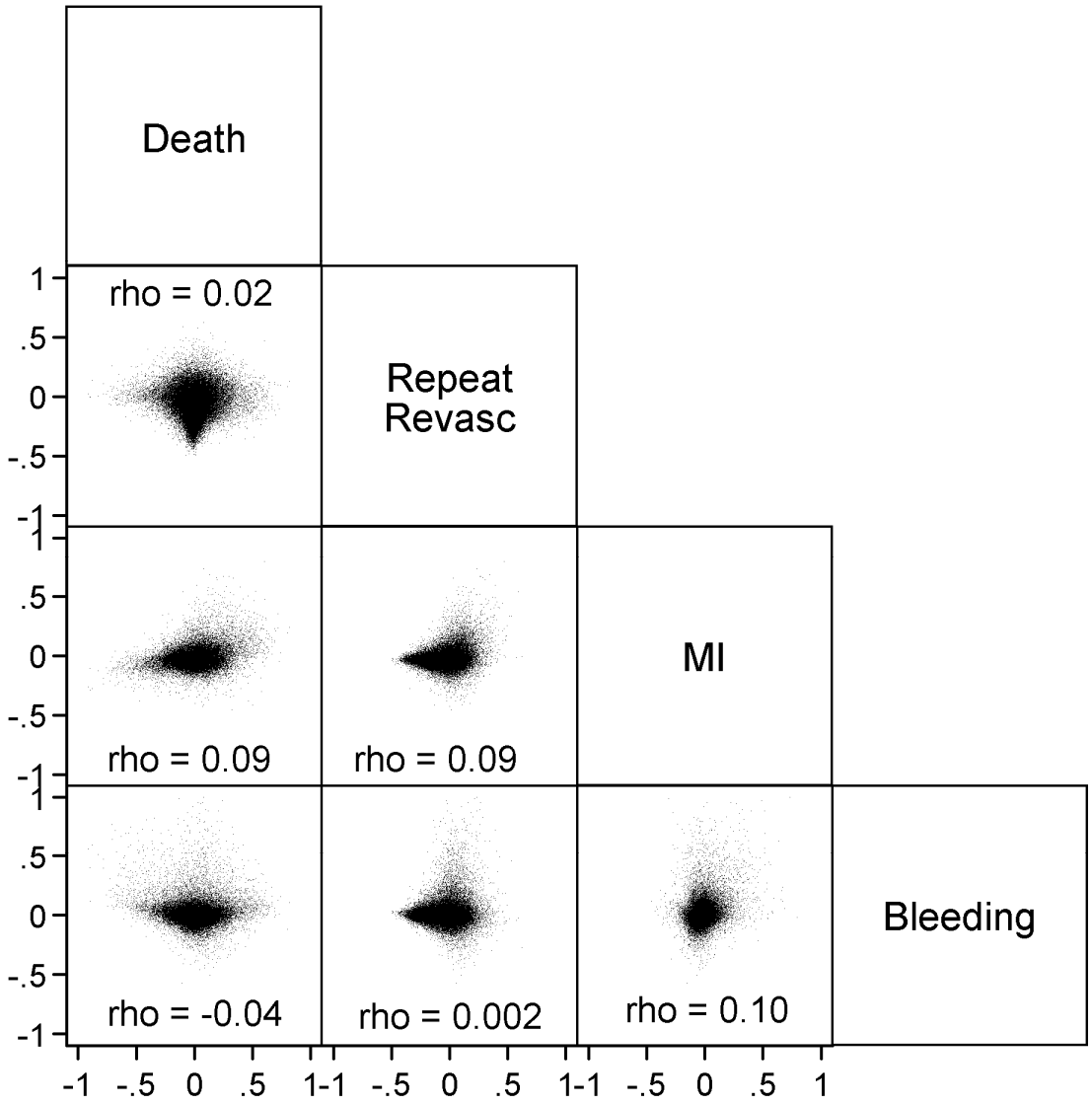


Figure 5.8: Correlations between individual-level PeT effects for different clinical outcomes.

Scatter plots display correlation between individual-specific PeT effects (the predicted PeT effects for each person in the cohort) across the different clinical outcomes assessed. Correlation was measured with a Spearman coefficient (rho).



6. CONCLUSIONS

STUDIES 1 AND 2

Summary of findings

In Study 1, we sought to determine the prevalence of stress testing in the years following elective PCI, both first post-PCI tests and additional post-PCI testing, predictors of stress testing after elective PCI, correlations between facility-level rates of stress testing use and baseline patient risk, and correlations between facility-level rates of stress testing use and subsequent clinical outcomes. We found that stress testing was commonly performed after elective PCI, with 58.2% of patients having at least one stress test by 27 months after PCI. Furthermore, patients who received one post-PCI stress test appeared to fall into a pattern of care in which they received additional stress tests on an annual basis subsequent to their first; a pattern highly suggestive of surveillance testing. Factors predicted to increase the risk of recurrent ischemia were generally paradoxically associated with lower rates of stress test utilization, even after adjustment for other clinical factors. Among a subset of higher risk patients, there was broad – over three-fold – variation in the rate of stress test utilization, based on the facility in which patients received PCI (a proxy for local health system-level practice patterns). However, facilities where patients received more stress tests after PCI were those in which patients were on average at lower risk of death, myocardial infarction, or repeat revascularization based on their observed characteristics than those treated in low test-use facilities. Correlations between facility-level rates of stress testing and clinical outcomes

were not significant for death or myocardial infarction, but higher rates of stress testing were correlated with higher rates of repeat revascularization.

In Study 2, we sought to evaluate associations between the imaging modality used to perform exercise stress testing after PCI (either echocardiography or nuclear imaging) and subsequent outcomes and resource use. We used an inverse probability weighted framework to adjust for potential confounding (endogenous treatment choice). We found that after adjustment, patients tested using echocardiography had lower rates of subsequent invasive cardiac procedures (catheterization or coronary revascularization) in the first 90 days after their stress test, but higher rates of repeat stress testing. Rates of both invasive cardiac procedure and repeat stress testing did not differ for events subsequent to 90 days. Total Medicare payments were lower in patients tested with echocardiography, but this difference was not significant after 14 months post-testing. Rates of death and MI-related readmission did not differ. Together, these findings suggest that the choice of imaging modality has a short term impact on processes of care, but that these effects offset each other so that there is a tradeoff between increased use of subsequent invasive testing and subsequent use of repeat stress testing. Cost differences between echocardiography and nuclear imaging appear driven primarily by differences in test reimbursement.

Limitations

The merged CathPCI-Medicare dataset used in Study 1 of this dissertation is the most detailed dataset of its kind, but it is not without limitations. Because the primary purpose of the CathPCI Registry is to improve the quality of PCI care, some data elements that would be of particular value for outcomes research are lacking. Socioeconomic status is unquestionably

an important factor in determining patient treatment choices and outcomes, even when individuals all have the same primary insurer (Medicare in this care). However, there are no socioeconomic data captured in the Registry, forcing reliance on ZIP code level data. ZIP code-level data are a notoriously error-prone approach to addressing individual socioeconomic status (168). We also do not have data on supplemental Medicare coverage, such as plans offered by a former employer or Medigap plans, so consequently it is not possible to determine whether differences in the cost of testing affect decision-making.

While data quality in CathPCI is validated with an onsite audit program, some have noted deficiencies with the data, both due to the far lower staffing levels for data collection when compared with randomized trials and the use of the patient's medical records as the "gold standard" by which the accuracy of registry data elements is assessed. The quality of data in the clinical record has been demonstrated using a single facility's experience to be remarkably poor (48), a finding which raises concern both for the validity of this study as well as for the ultimate value of all registries derived from clinical data for conducting observational CER.

While Study 2's dataset was also lacking socioeconomic data, it was even more limited by virtue of not including the CathPCI data elements and thus being entirely reliant on claims data. As the primary purpose of claims is to facilitate payment for services rather than to document the patient's condition when services were received, they are far from an optimal information source for CER. Relying on the quality of data recorded in billing records to identify all sources of potential confounding due to clinical factors is known to be problematic (169). While large scale registries clearly have problems, as outlined above, the

use of the combined registry-claims datasets has been shown to improve model prediction capability (170).

These studies also have several limitations in common. For both studies, the only data available at time of stress testing (if any) was that which could be gleaned from claims records. This limitation precludes identifying test indication and patient condition at time of testing. Both of these studies rely on data from the Federal Medicare program. Because of the limited Medicare eligibility for adults younger than age 65 (e.g., receiving Social Security disability benefits or having been diagnosed with end stage renal disease), the cohorts were limited to older Americans, and study results do not necessarily generalize to younger Americans. Like all studies utilizing Fee-for-Service Medicare claims, both studies outlined here are potentially vulnerable to threats to their external generalizability by the fact that Medicare Advantage recipients were not observed. Medicare Advantage recipients are generally younger and healthier than Fee-for-Service beneficiaries (171). Future work could attempt to adjust for this problem either by jointly modeling Fee-for-Service selection (i.e., a Heckman selection model), or by using inverse probability of weighting approaches from the epidemiology literature (weights for various purposes can be multiplied together) (172).

The patients studied here received treatment between 2003-2008. Cardiology is a rapidly evolving medical specialty, and it is possible that other changes in cardiology care have changed the population receiving PCI, or the population receiving stress testing after PCI, in ways that threaten the applicability of these results to modern practice.

Policy implications and future directions

Study 1 demonstrates that if anything, patients who receive stress testing after PCI are a population expected to receive a *lower* benefit from such testing than the average PCI recipient. Thus, policy makers may be confident based on these results that the disconnect between the guidelines consensus and clinical practice is not due to physicians employing a personalized approach based on patient risk rather than following broad guidelines. Furthermore, even in a higher risk population that might be expected to receive greater than average benefit from surveillance testing, facility-level stress testing rates were correlated with higher rates of repeat revascularization, but no improvement in rates of death or myocardial infarction. While relying on indirect, observational evidence is obviously not preferred, these results do suggest that current guidelines considering the use of routine surveillance stress testing after PCI to be inappropriate are correct.

Study 2 demonstrates that the choice of echocardiography versus nuclear imaging results in a modest, short term difference in processes of care in which echocardiography patients receive additional stress tests, but reduced rates of catheterization and repeat revascularization. Most of the difference in the cost of care appears driven by differences in the reimbursement for the tests themselves. Given the modest differences noted here, it appears that leaving the choice in the hands of patients and providers, based on preferences and local familiarity with both techniques, is the most reasonable approach. For future research, it would be important to understand if the regional variation in preference for echocardiography versus nuclear imaging results in productivity spillovers (173). If so, allowing for continued variation in practice patterns is likely to improve aggregate welfare.

Future research concerning the optimal use of stress testing after PCI would be enhanced by the development of a prospective registry, with data at the time of stress testing rather than at the time of PCI. Experience with the CathPCI and other NCDR registries, as well as similar quality improvement initiatives such as the American Heart Association's "Get with the Guidelines" project has led to the development of automated data collection tools that transfer standardized data elements from popular electronic medical records platforms directly into the registry case report form, reducing the cost of registry participation. Furthermore, the development of a stress testing registry would lower the cost of future studies by providing a common infrastructure through which standardized data elements can be rapidly captured and study-specific data elements prospectively incorporated with minimal incremental effort for participating sites. Such efforts have already begun with CathPCI, as it is now the backbone of an observational pharmaceutical post-marketing safety study (TRANSLATE-ACS)(174) as well as a randomized trial of PCI access techniques in women (SAFE-PCI).

Ultimately, if additional research continues to demonstrate the validity of the AUC, one approach to controlling costs may be for insurers to base their coverage decisions directly on the AUC, as the Delaware Blue Cross Blue Shield plan was forced to adopt after a state investigation into their imaging review practices (175). Doing so would replace a system in which individual insurance companies make often-conflicting coverage decisions based on the expertise of their medical reviewers with a consistent, consensus-driven and evidence-based approach.

STUDY 3

Summary of findings

In Study 3, we sought to demonstrate the use of a recently-developed econometric technique - local instrumental variables - for estimating policy relevant estimands in outcomes research. We applied the methodology to the choice of coronary stent for patients undergoing PCI - either a drug-eluting stent or a bare metal stent. We find evidence of considerable endogeneity in treatment selection that is not completely removed with conventional regression methods. In the local instrumental variables framework, we found evidence that the safety and efficacy profile of drug-eluting stents was considerably more favorable among the patients who actually received them than those that did not.

Limitations

The Medicare dataset used in this Aim was similar to that which was employed in Study 2. As with Study 2, the lack of some data elements does call into question the completeness of adjustment for potential confounding between DES and BMS. While a notable limitation, in our other studies which has used the CathPCI Registry with linked Medicare data, results have been consistent with what is reported in this dissertation. The choice of drug-eluting versus bare metal stenting likely depends, at least in part, on how affordable clopidogrel is for the patient, as patients who discontinue clopidogrel while on DES face dramatically higher risk of stent thrombosis and subsequent death (176).

Policy implications and future directions

Study 3 highlights the importance of a thoughtful approach to patient selection for DES insertion, and the potential benefits of local instrumental variables methods for outcomes research:

1. Our results suggest that in appropriately selected patients, DES are safe and efficacious, while in poorly selected patients, they are less effective and may cause harm. These results should reinforce to interventional cardiologists the importance of careful patient selection for DES placement. One possible further line of research that may assist in this effort would be to develop more clinically-nuanced models of predicted DES safety and efficacy that could interface with electronic medical records platforms to automatically estimate patient-specific predictions. In the meantime, DES rates should be monitored by policymakers. Rates that approach the >90% utilization rate observed in 2005 ought to be cause for concern, unless it is demonstrated that the safety and efficacy profile of newer DES, with more modern antiplatelet agents and treatment paradigms has altered the risk-benefit ratio.

2. In order for methods such as local instrumental variables to meet their potential, a great deal of additional methodological research is needed to understand issues such as: a) a tractable approach to model fitting in the setting of a complicated multi-stage equation, particularly with highly dimensional data; b) a theoretical and practical understanding of the assumptions underlying such models, and the consequences of violations; c) "best practices" for a structured approach to regression implementation, in the same manner as Peter Austin and colleagues have developed for propensity score methods; and d) an understanding of why bias occurs in bootstrapped replicates. An important limitation of LIV is that effects are

only identified when the region of common support for the propensity score fills close to the entire interval of (0,1), where the definition of "close" has never been formally defined (159). Obtaining a propensity score with that much range in values requires a strong instrument and/or covariates that strongly affect treatment receipt; while it was possible to obtain a fairly broad range (0.09-0.81) in this study, whether such range in obtainable in most empirical work or adequate for analysis is unclear. In summary, LIV is a methodology with tremendous promise, but for which much additional work is needed to assess its practical value and to provide guidance on its implementation.

SUMMARY

Improvements in cardiology care improve the quality and length of human life, and observational methods have an important role to play in ensuring that cardiovascular care continues to improve in quality, but at an acceptable financial cost. In this dissertation, I produced three papers highlighting applications of observational methods to questions concerning the management of patients receiving and who have received PCI. The results from these studies will inform clinical practice for the care of PCI patients as well as paving the way for additional improvements in how outcomes research is performed.

APPENDIX A: DERIVATION OF INITIAL STUDY SAMPLE

This section is adapted from documentation provided by John Cantrell, based on programming work performed by John Cantrell.

Phase 1

We requested all Medicare Provider Analysis and Review (MedPAR) claims for 2003-2004 with a diagnosis code of (410.xx (myocardial infarction), 411.1 (unstable angina), or 403.9 (angina NOS)). From these claims, the first record per subject was considered the index hospitalization. Subjects were retained whose index hospitalizations were not:

1. admissions to a facility other than a short stay or critical access hospital;
2. did not included diagnosis codes indicative of history of coronary bypass or PCI (V45.81-V45.82, 414.02-414.06, 996.72);
3. did not include receipt of valve replacement (procedure codes 35.xx).

These exclusions produced a population of **1.29M** subjects

Phase 2

For the 1.29M subjects identified above, 2002-2006 Medicare denominator files were obtained. For each subject, a start date was identified 12 months prior to the admission date of their index hospitalization, or January 1, 2002 for patients whose admission date antedated January 1, 2003. In addition, an end date was identified as being their data of death or December 31, 2006; whichever occurred later. Subjects were removed from the cohort if they experienced any of the following between their study entry or exit dates:

1. Multiple denominator datasets in single year file

2. Reported location outside of the 50 US States, plus the District of Columbia.

3. Disenrollment from either Medicare Part A or B

4. Age < 65 at study start date

5. Enrollment in a Medicare Advantage plan

Note: Because of this exclusion (which is not modifiable), we do not have data on patients who enrolled in Medicare Advantage subsequent to their index event. Thus, we have an "always enrolled" cohort for which the only source of censoring is due to different length of follow-up based on when in 2003-2004 they experienced their index event.

These exclusions produced a population of **1,096,614** subjects.

Phase 3

Per the terms of our contract with the Centers for Medicare and Medicaid Services (CMS), we were permitted to obtain all 2002-2008 MedPAR (not just those with a diagnosis code for ACS/angina), Carrier, Outpatient, and Denominator files for 1 million subjects. Thus, we needed to remove 96,614 patients from the request. This removal was performed by eliminating those with the most recent admission dates for their index hospitalization (i.e., starting at December 31, 2004) and working backwards. 1,000,000 subjects was reached excluding all subjects admitted 10/15/2004-12/31/2004, and some subjects admitted 10/14/2004. Subsequent to the data request being sent to the CMS contractor, a coding error was discovered that reduced the ultimate sample size to **987,860**.

APPENDIX B: R CODE TO IMPLEMENT PET-BASED IMPLEMENTATION OF LOCAL IV

```
# ************************************************************************
# Program Name  : 03_petmodel.R
# Project       : Local IV Project
# Description   : Reimplements PET effects using Basu methodology
# Programmer    : Jeff Federspiel (jerome_federspiel@med.unc.edu)
# Original Date : 10/20/2012 (M/D/Y)
# Input Files   : desbms_ahj_done
# Output Files  : bootout.dta
# R Command     : R CMD BATCH 06_petmodel.R
# ************************************************************************

# *********** Startup tasks ************************************************

options(width=9999)
cat("Program started", date())

# Clear memory
rm(list=ls())
gc()

# Change working directory
setwd("/hpm2/acs/jeff/desbmsliv/logs")

# Load required libraries (use install.packages() if any are missing)
library(splines)
library(stats)
library(foreign)
library(parallel)
library(gdata)
library(gtools)

# *********** Data Load ****************************************************

# Data load
indat <- read.dta("../sasdsl/desbms_ahj_cleaned.dta", convert.factors=FALSE)

dim(indat)

# *********** Regression Parameters ****************************************

# Covariates to be included in model
termlist <- NULL

intlist <- c("age_70","age_75","age_80","age_85","age_85up","nonwhite","female",
             "inc_100","inc_200","inc_300","inc_400","inc_500up", "dual_elig",
             "micropolitan", "smalltown","rural", "medsch",
             "forprofit","govt","midwest","northeast","west",
             "stemi","nstemi","pci_multv",
             "chf","valve","pulmcirc","perivasc","para","neuro",
             "chrnlung","diabetes","hypothy","renlfail","cancer",
             "arth","obese","wghtloss","lytes","anemia","psych",
             "depress","htn_c")

# List of outcome variables to use
outcomelist <- c("died46m","revasc46m","ami46m","bleed46m")

# For each outcome, regression model to use
functionlist <- c("logit","logit","logit","logit")

# For each outcome, degree of polynomial of propensity score to use in second stage
degreelist <- c(1,1,1,1)

# Number of points to sample for distribution of u
simsize <- 100

# Number of replicates to run for bootstrap
```

```
R <- 500

# Number of CPU cores to use
cores <- 8

# Type of approximation of U to employ
umethod <- 2

# Binary (treatment) variable
treatvar <- "desonly"

# ********** PET Function ***************************************************

# Main function to produce all average effects
jeffpet <- function(data, treatvar, termlist=NULL, intlist=NULL, outcomes, functions,
                    simsize, degrees, facvar, umethod=1, display=TRUE, personid)
{

  if (all(na.omit(data[[treatvar]])) %in% 0:1) {

  }
  else {
        cat(treatvar, " must be 0/1 binary")
        return()
  }

  fmlterms <- as.formula(paste(paste(treatvar)," ~ (", paste(c(termlist, intlist),
collapse="+"),
                         " ) + as.factor(revascmonth)"))

  psmodel <- glm(fmlterms,
                 data=data,
                 family=binomial(link=logit))


  if (display==TRUE) {
    cat("\n###############################################################################")
    cat("\nPropensity Score Model\n")
    print(summary(psmodel))
  }

  # Add on first stage value to dataset
  data$ps <- psmodel$fitted.values
  rm(psmodel)

  # Trim outside common support
  minps  <- max(min(subset(data$ps, data[[treatvar]]==TRUE)),
                min(subset(data$ps, data[[treatvar]]==FALSE)))

  maxps  <- min(max(subset(data$ps, data[[treatvar]]==TRUE)),
                max(subset(data$ps, data[[treatvar]]==FALSE)))


  orig <- dim(data)[1]
  data <- subset(data, (minps <= data$ps) & (data$ps <= maxps))

  # Display sample size in trimmed sample
  if (display==TRUE) {
    cat(paste("Common support from ", round(minps,3),
              " to ", round(maxps,3)), "\n",
              paste("N with common support = ", dim(data)[1]), " of ", orig, "
(",format((dim(data)[1]*100)/orig, digits=4),"%)", "\n",sep="")
  }

  # Calculate sample size in sample
  N <- dim(data)[1]

  # Initialize output matrix for results
  output <- array(0, c(length(outcomes), 4))
  rownames(output) <- outcomes
  colnames(output) <- c("ate", "tt", "tut", "tt_tut")
```

```
  # Initialize output dataset for PETs
  pets <- array(NA, c(N, length(outcomes)))
  colnames(pets) <- c(outcomelist)

  pets <- as.data.frame(pets)

  # For each outcome, run second stage model and calculate effects
  for(i in 1:length(outcomes)) {

    # Set up formula and run model
    if (degrees[i] == 1) { ppart <- NULL }
    if (degrees[i] == 2) { ppart <- " + I(ps^2)" }
    if (degrees[i] == 3) { ppart <- " + I(ps^2) + I(ps^3)" }
    if (degrees[i] == 4) { ppart <- " + I(ps^2) + I(ps^3) + I(ps^4)" }
    if (length(intlist)>0) {
           formula <- as.formula(paste(outcomes[i], " ~ ", paste(termlist, collapse="+"), " +
(", paste(intlist, collapse="+"), " ) * ps", ppart))
    }
    else {
           formula <- as.formula(paste(outcomes[i], " ~ ", paste(termlist, collapse="+"),"+
ps", ppart))
    }


    model <- glm(formula,
                 data=data,
                 family=binomial(link=functions[i]))

    if (display==TRUE) {
      cat("\n###############################################################################")
      cat(paste("\n",paste("Outcome =",outcomes[i]),"\n",sep=""))
      print(summary(model))
    }

    # Extract coefficients for main effects, pscore terms, and interaction terms
    mainterms <- model$coef[1:(length(c(termlist,intlist))+1)] # Includes constant term

    psterm  <- model$coef[(length(c(termlist,intlist))+2)]

    psterm2 <- 0
    psterm3 <- 0
    psterm4 <- 0

    if (degrees[i]>1) { psterm2 <- model$coef[(length(c(termlist,intlist))+3)] }
    if (degrees[i]>2) { psterm3 <- model$coef[(length(c(termlist,intlist))+4)] }
    if (degrees[i]>3) { psterm4 <- model$coef[(length(c(termlist,intlist))+5)] }

    # Interaction terms
    if (length(intlist) > 0) {
      intterms  <-
model$coef[(length(c(termlist,intlist))+2+degrees[i]):(length(intlist)+length(c(termlist,intl
ist))+1+degrees[i])]
    }

    else {
     intterms <- matrix(0,N,1)
    }

    rm(model)

    # Replace any omitted terms with zeros
    intterms[is.na(intterms)] <- 0
    mainterms[is.na(mainterms)] <- 0
    psterm[is.na(psterm)] <- 0
    psterm2[is.na(psterm2)] <- 0
    psterm3[is.na(psterm3)] <- 0
    psterm4[is.na(psterm4)] <- 0

    # Get score matrices (XB for interactions with p-score,
    # and intercept plus main effects)
```

```
    intscore <- as.matrix(data[,c(intlist)]) %*% as.matrix(intterms)
    mainscore <- as.matrix(cbind(1, data[,c(termlist, intlist)])) %*% as.matrix(mainterms)
    rm(intterms, mainterms)

    # Several different potential formulas for u
    # 1. From min to max propensity score ala Basu using random draws
    if (umethod==1) { u = matrix(runif(N*simsize, min(data$ps), max(data$ps)), N, simsize) }

    # 2. From min to max propensity score (deterministically)
    if (umethod==2) { u = matrix(seq(minps, maxps, length=simsize), N, simsize, byrow=T) }

    # 3. From nearly 0 to nearly 1 (deterministically)
    if (umethod==3) { u = matrix(seq((1/simsize)/2, (simsize-1)/simsize+(1/simsize)/2,
length=simsize), N, simsize, byrow=T) }

    # Calculate XB for each individual at each value of u
    xbhat <- matrix(mainscore,N,simsize,byrow=F) + # X part
             matrix(intscore,N,simsize,byrow=F)*u + # X*ps part
             psterm*u + psterm2*(u^2) + psterm3*(u^3) + psterm4*(u^4) # K(ps) part

    # Caculate MTE (partial derivative with respect to pscore term)
    if (functions[i] == "logit") {
      mte = inv.logit(xbhat) * (1-inv.logit(xbhat)) *
            (matrix(intscore, N, simsize, byrow=F) + psterm + 2*psterm2*u + 3*psterm3*(u^2) +
4*psterm4*(u^3))
    }

    if (functions[i] == "probit") {
      mte = dnorm(xbhat) *
            (matrix(intscore, N, simsize, byrow=F) + psterm + 2*psterm2*u + 3*psterm3*(u^2) +
4*psterm4*(u^3))
    }

    # Replace MTEs with absolute values > 1 with 1 (Done in Basu Mata code)
    mte[abs(mte)>1] <- sign(mte[abs(mte)>1])*1

    # Remove unneeded xbhat matrix
    rm(xbhat)
    gc()

    # Generate indicator of whether ps > u
    dstar1 <- (qnorm(data$ps) + qnorm(1-u)) > 0

    # Remove unneed u matrix
    rm(u)
    gc()

    # TT is average MTE for values of ps > u
    tt <- as.matrix(rowSums(mte * dstar1)/rowSums(dstar1))

    # TUT is average MTE for values of ps < u
    tut <- as.matrix(rowSums(mte * (1-dstar1))/rowSums(1-dstar1))

    # Clean-up unneeded MTE and dstar1 matrices
    rm(mte, dstar1)
    gc()

    # PET effect is TT for treated patients and TUT for untreated patients
    pet       <- tt*(data[[treatvar]]==TRUE) + tut*(data[[treatvar]]==FALSE)
    pets[i]   <- pet # Outputs the pet effects to results matrix

    # ATE is overall average
    ate       <- mean(pet, na.rm=TRUE)

    # TT (average PET among Treated recipients)
    att       <- mean(subset(pet, data[[treatvar]]==TRUE), na.rm=TRUE)

    # TUT (PET among Untreated recipients)
    atut       <- mean(subset(pet, data[[treatvar]] ==FALSE), na.rm=TRUE)

    # Save results for ATE, TT, TUT, TT_TUT to output matrix
```

```
    output[i,] <- c(ate, att, atut, att-atut)

    # Clean up unneeded datasets
    rm(tt, tut, pet, ate, att, atut,
       intscore, mainscore,
       psterm, psterm2, psterm3, psterm4)
    gc()

  }

  # Clean up final datasets
  pets <- as.data.frame(cbind(data[[personid]], data[[treatvar]], pets, data$ps))
  rm(data, i)
  gc()

  # Output results
  return(list(output=output, pets=pets))
  rm(output, pets)
  gc()
}

# *********** Base Case ***************************************************

base <- jeffpet(data=indat,              # Input data
        treatvar=treatvar,               # Treatment variable
        termlist=termlist,               # Variables to include
        intlist=intlist,                 # Variables to include with interactions
        outcomes=outcomelist,            # Outcomes to assess
        functions=functionlist,          # Models to use
        degrees=degreelist,              # Polynomial of pscore to use
        facvar="m_revaschospprovno",     # Facility ID variable
        personid = "bid",                # Person identifier
        simsize=simsize,                 # Number of points to use to approximate u
        umethod = umethod,               # Method to create values of u
        display=TRUE)                    # Display output of regressions

base$output

# Renaming variables b/c Stata chokes on variable names with $ symbols in them
base$pets <- rename.vars(base$pets, from=c("data[[personid]]", "data[[treatvar]]",
"data$ps"), to=c("bid","anydes","ps"), info=FALSE)

# *********** Bootstrapping ***********************************************

# Function for clustered bootstrap sampling
# Modified from http://biostat.mc.vanderbilt.edu/wiki/Main/HowToBootstrapCorrelatedData

resample <- function(dat, cluster, replace, addid) {

  # exit early for trivial data
  if(nrow(dat) == 1 || all(replace==FALSE))
      return(dat)

  # sample the clustering factor
  cls <- sample(unique(dat[[cluster[1]]]), replace=replace[1])

  # subset on the sampled clustering factors
  # Add new identifying variable (sampleid) for highest-level
  if (addid == TRUE) {
    sub <- lapply(seq(1:length(cls)),
                  function(sampleid)
                  cbind(subset(dat, dat[[cluster[1]]]==cls[sampleid]),
                        sampleid))
  }

  # Don't add identifying variable for subsequent levels
  if (addid == FALSE) {
    sub <- lapply(cls, function(b) subset(dat, dat[[cluster[1]]]==b))
  }

  # sample lower levels of hierarchy (if any)
```

```
  if(length(cluster) > 1) sub <- lapply(sub, resample, cluster=cluster[-1], replace=replace[-
1],addid==FALSE)

  # join and return samples
  do.call(rbind, sub)

}

# Wrapper function for each bootstrap cycle
bootwrapper <- function(index) {

  repset <- indat[sample(nrow(indat), nrow(indat), replace=T),]

  bootfxn <- jeffpet(data=repset,                    # Input data
                     treatvar=treatvar,              # Treatment variable
                     termlist=termlist,              # Variables to include
                     intlist=intlist,                # Variables to include with interactions
                     outcomes=outcomelist,           # Outcomes to assess
                     functions=functionlist,         # Models to use
                     degrees=degreelist,             # Polynomial of pscore to use
                     facvar="sampleid",              # Facility ID variable
                     personid="bid",                 # Person ID variable
                     simsize=simsize,                # Number of points to approximate u
                     umethod=umethod,                # Method to approximate distriubiton of u
                     display=FALSE)                  # Suppress display

  output <- as.vector(bootfxn$output)
  return(output)
  rm(repset, bootfxn, output)
  gc()
}

# Do the actual bootstrap
starttime <- proc.time()

set.seed(19075020)
bootresults <- mclapply(1:R,
                        bootwrapper,
                        mc.set.seed=TRUE,
                        mc.silent=TRUE,
                        mc.cores=cores,
                        mc.cleanup=TRUE)
proc.time()-starttime

# Function to get bias-corrected CIs
cifunction <- function(boot.out,base.out,conf=0.95)
{
  dataset <- as.matrix(do.call(rbind,boot.out))

  correction <- qnorm(colMeans(dataset < matrix(base.out, R, length(base.out), byrow=T)))

  ll <- sapply(seq(1:dim(dataset)[2]),
               function(value) {quantile(dataset[,value],
pnorm(qnorm(0.025)+2*correction[value]))})

  ul <- sapply(seq(1:dim(dataset)[2]),
               function(value) {quantile(dataset[,value],
pnorm(qnorm(0.975)+2*correction[value]))})

  matrixcis <- cbind(
    seq(1:length(as.vector(base.out))), # Sequential Number
    as.vector(base.out), # Point estimate
    as.vector(ll), # Lower bound
    as.vector(ul), # Upper bound
    colMeans(dataset)-as.vector(base.out)) # Bias estimate (difference between mean value of
bootstraps and point estimate)

  colnames(matrixcis) <- c("index","point","ll","ul","bias")

  rownames(matrixcis) <- sapply(seq(1:length(as.vector(base.out))),
                                function(index) {
```

```
                                paste(colnames(base.out)[ceiling(index/length(outcomelist))],
                                rownames(base.out)[(index-1)%%length(outcomelist)+1],
sep="_")
                                })

  colnames(dataset) <- rownames(matrixcis)

  return(list("Summary"=matrixcis,"Boot_Iterations"=dataset))
}

bootout <- cifunction(bootresults, base$output)
bootout$Summary

# Save output
save(bootout, file="../output/bootout_int.rData")
write.dta(base$pets, "../output/pets.dta")
write.dta(as.data.frame(bootout$Boot_Iterations), "../output/bootout_iterations.dta")
write.dta(as.data.frame(bootout$Summary), "../output/pet_summary.dta")

# ****************** Wrapup ************************************************

# Empty workspace
rm(list=ls())

cat("Program ended", date())

# ****************** End of 06_petmodel.R **********************************
```

BIBLIOGRAPHY

1. National Heart Lung and Blood Institute. Morbidity & Mortality: 2009 Chart Book on Cardiovascular, Lung, and Blood Diseases. Bethesda: National Institutes of Health.

2. Wijeysundera HC, Machado M, Farahati F, Wang X, Witteman W, van der Velde G, et al. Association of temporal trends in risk factors and treatment uptake with coronary heart disease mortality, 1994-2005. JAMA. 2010;303(18):1841-7.

3. Hunink MG, Goldman L, Tosteson AN, Mittleman MA, Goldman PA, Williams LW, et al. The recent decline in mortality from coronary heart disease, 1980-1990. The effect of secular trends in risk factors and treatment. JAMA. 1997 Feb 19;277(7):535-42.

4. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. N Engl J Med. 2007 Jun 7;356(23):2388-98.

5. Yusuf S, Zhao F, Mehta SR, Chrolavicius S, Tognoni G, Fox KK. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. N Engl J Med. 2001 Aug 16;345(7):494-502.

6. Rose EA, Gelijns AC, Moskowitz AJ, Heitjan DF, Stevenson LW, Dembitsky W, et al. Long-term use of a left ventricular assist device for end-stage heart failure. N Engl J Med. 2001 Nov 15;345(20):1435-43.

7. Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, De Marco T, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. N Engl J Med. 2004 May 20;350(21):2140-50.

8. Krumholz HM, Wang Y, Chen J, Drye EE, Spertus JA, Ross JS, et al. Reduction in acute myocardial infarction mortality in the United States: risk-standardized mortality rates from 1995-2006. JAMA. 2009 Aug 19;302(7):767-73.

9. Krumholz HM, Radford MJ, Ellerbeck EF, Hennen J, Meehan TP, Petrillo M, et al. Aspirin in the treatment of acute myocardial infarction in elderly Medicare beneficiaries. Patterns of use and outcomes. Circulation. 1995 Nov 15;92(10):2841-7.

10. Masoudi FA, Foody JM, Havranek EP, Wang Y, Radford MJ, Allman RM, et al. Trends in acute myocardial infarction in 4 US states between 1992 and 2001: clinical characteristics, quality of care, and outcomes. Circulation. 2006 Dec 19;114(25):2806-14.

11. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. Lancet. 2003 Jan 4;361(9351):13-20.

12. Jollis JG, Roettig ML, Aluko AO, Anstrom KJ, Applegate RJ, Babb JD, et al. Implementation of a statewide system for coronary reperfusion for ST-segment elevation myocardial infarction. JAMA. 2007 Nov 28;298(20):2371-80.

13. Le May MR, So DY, Dionne R, Glover CA, Froeschl MP, Wells GA, et al. A citywide protocol for primary PCI in ST-segment elevation myocardial infarction. N Engl J Med. 2008 Jan 17;358(3):231-40.

14. Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. Circulation. 2011 Mar 1;123(8):933-44.

15. Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, et al. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010 Feb 23;121(7):e46-e215.

16. Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds Boards of Trustees. The 2011 annual report of the boards of trustees of the federal hospital insurance and federal supplementary medical insurance trust funds. Washington, DC: 2011.

17. Fleurence RL, Naci H, Jansen JP. The critical role of observational evidence in comparative effectiveness research. Health Aff (Millwood). 2010 Oct;29(10):1826-33.

18. Heiat A, Gross CP, Krumholz HM. Representation of the elderly, women, and minorities in heart failure clinical trials. Arch Intern Med. 2002 Aug 12-26;162(15):1682-8.

19. Dunn KM, Jordan K, Lacey RJ, Shapley M, Jinks C. Patterns of consent in epidemiologic research: evidence from over 25,000 responders. Am J Epidemiol. 2004 Jun 1;159(11):1087-94.

20. Williams B, Irvine L, McGinnis AR, McMurdo ME, Crombie IK. When "no" might not quite mean "no"; the importance of informed and meaningful non-consent: results from a survey of individuals refusing participation in a health-related research project. BMC Health Serv Res. 2007;7:59.

21. Andrus BW, Welch HG. Medicare services provided by cardiologists in the United States: 1999-2008. Circ Cardiovasc Qual Outcomes. 2012 Jan 10.

22. Lucas FL, DeLorenzo MA, Siewers AE, Wennberg DE. Temporal trends in the utilization of diagnostic testing and treatments for cardiovascular disease in the United States, 1993-2001. Circulation. 2006 Jan 24;113(3):374-9.

23. United States Government Accountability Office. GAO-08-452: Medicare Part B imaging services: rapid spending growth and shift to physician offices indicate need for CMS to consider additional management practices. Washington 2008.

24. Levin DC, Parker L, Intenzo CM, Sunshine JH. Recent rapid increase in utilization of radionuclide myocardial perfusion imaging and related procedures: 1996-1998 practice patterns. Radiology. 2002 Jan;222(1):144-8.

25. Center for the Evaluative Clinical Sciences at Dartmouth Medical School. The Dartmouth atlas of cardiovascular health care. Chicago: AHA Press; 1999.

26. Levin DC, Intenzo CM, Rao VM, Frangos AJ, Parker L, Sunshine JH. Comparison of recent utilization trends in radionuclide myocardial perfusion imaging among radiologists and cardiologists. J Am Coll Radiol. 2005 Oct;2(10):821-4.

27. Levin DC, Rao VM, Parker L, Frangos AJ, Intenzo CM. Recent payment and utilization trends in radionuclide myocardial perfusion imaging: comparison between self-referral and referral to radiologists. J Am Coll Radiol. 2009 Jun;6(6):437-41.

28. Shah BR, Cowper PA, O'Brien SM, Jensen N, Patel MR, Douglas PS, et al. Association between physician billing and cardiac stress testing patterns following coronary revascularization. JAMA. 2011 Nov 9;306(18):1993-2000.

29. Douglas PS, Taylor A, Bild D, Bonow R, Greenland P, Lauer M, et al. Outcomes research in cardiovascular imaging: report of a workshop sponsored by the National Heart, Lung, and Blood Institute. JACC Cardiovasc Imaging. 2009 Jul;2(7):897-907.

30. Wennberg DE, Kellett MA, Dickens JD, Malenka DJ, Keilson LM, Keller RB. The association between local diagnostic testing intensity and invasive cardiac procedures. JAMA. 1996 Apr 17;275(15):1161-4.

31. Centers for Medicare and Medicaid Services. 2010 Medicare Physician Fee Schedule. Washington: United States Department of Health and Human Services.

32. Hendel RC. Utilization management of cardiovascular imaging pre-certification and appropriateness. JACC Cardiovasc Imaging. 2008 Mar;1(2):241-8.

33. Mitchell JM, Lagalia RR. Controlling the escalating use of advanced imaging: the role of radiology benefit management programs. Med Care Res Rev. 2009 Jun;66(3):339-51.

34. Bonow RO, Douglas PS, Buxton AE, Cohen DJ, Curtis JP, Delong E, et al. ACCF/AHA methodology for the development of quality measures for cardiovascular technology: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures. Circulation. 2011 Sep 27;124(13):1483-502.

35. Shaw LJ, Marwick TH, Zoghbi WA, Hundley WG, Kramer CM, Achenbach S, et al. Why all the focus on cardiac imaging? JACC Cardiovasc Imaging. 2010 Jul;3(7):789-94.

36. Curtis LH, Greiner MA, Shea AM, Whellan DJ, Hammill BG, Schulman KA, et al. Assessment of Left Ventricular Function in Older Medicare Beneficiaries With Newly Diagnosed Heart Failure. Circ Cardiovasc Qual Outcomes. 2010 Nov 23.

37. Douglas PS, Garcia MJ, Haines DE, Lai WW, Manning WJ, Patel AR, et al. ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 Appropriate Use Criteria for echocardiography. J Am Coll Cardiol. 2011 Mar 1;57(9):1126-66.

38. Hendel RC, Berman DS, Di Carli MF, Heidenreich PA, Henkin RE, Pellikka PA, et al. ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 Appropriate use criteria for cardiac radionuclide imaging. J Am Coll Cardiol. 2009 Jun 9;53(23):2201-29.

39. Taylor AJ, Cerqueira M, Hodgson JM, Mark D, Min J, O'Gara P, et al. ACCF/SCCT/ACR/AHA/ASE/ASNC/NASCI/SCAI/SCMR 2010 appropriate use criteria for cardiac computed tomography. A report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the Society of Cardiovascular Computed Tomography, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the American Society of Nuclear Cardiology, the North American Society for Cardiovascular Imaging, the Society for Cardiovascular Angiography and Interventions, and the Society for Cardiovascular Magnetic Resonance. J Am Coll Cardiol. 2010 Nov 23;56(22):1864-94.

40. Douglas PS. Improving imaging: our professional imperative. J Am Coll Cardiol. 2006 Nov 21;48(10):2152-5.

41. Douglas PS, Chen J, Gillam L, Hendel R, Hundley WG, Masoudi F, et al. Achieving Quality in Cardiovascular Imaging II: proceedings from the Second American College of Cardiology -- Duke University Medical Center Think Tank on Quality in Cardiovascular Imaging. JACC Cardiovasc Imaging. 2009 Feb;2(2):231-40.

42. Douglas P, Iskandrian AE, Krumholz HM, Gillam L, Hendel R, Jollis J, et al. Achieving quality in cardiovascular imaging: proceedings from the American College of Cardiology-Duke University Medical Center Think Tank on Quality in Cardiovascular Imaging. J Am Coll Cardiol. 2006 Nov 21;48(10):2141-51.

43. Mudrick D, Kaltenbach LA, Shah B, Lytle B, Masoudi FA, Mark DB, et al. Downstream testing and subsequent procedures after coronary computed tomographic angiography following coronary stenting in patients >/=65 years of age. Am J Cardiol. 2012 Sep 15;110(6):776-83.

44. Mudrick D, Shah BR, McCoy LA, Lytle BL, Masoudi FA, Federspiel JJ, et al. Patterns of stress testing and diagnostic catheterization after coronary stenting in 250,350 Medicare beneficiaries. Circ Cardiovasc Imaging. 2012 Oct 16.

45. Douglas PS, Khandheria B, Stainback RF, Weissman NJ, Peterson ED, Hendel RC, et al. ACCF/ASE/ACEP/AHA/ASNC/SCAI/SCCT/SCMR 2008 appropriateness criteria for stress echocardiography: a report of the American College of Cardiology Foundation Appropriateness Criteria Task Force, American Society of Echocardiography, American College of Emergency Physicians, American Heart Association, American Society of Nuclear Cardiology, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance endorsed by the Heart Rhythm Society and the Society of Critical Care Medicine. J Am Coll Cardiol. 2008 Mar 18;51(11):1127-47.

46. Gibbons RJ, Balady GJ, Bricker JT, Chaitman BR, Fletcher GF, Froelicher VF, et al. ACC/AHA 2002 guideline update for exercise testing: summary article. A report of the

American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines). J Am Coll Cardiol. 2002 Oct 16;40(8):1531-40.

47. American College of Cardiology. Five things physicians and patients should question: American Board of Internal Medicine Foundation web site; 2012 [August 1, 2012]. Available from: http://choosingwisely.org/wp-content/uploads/2012/04/5things_12_factsheet_Amer_Coll_Cardio.pdf.

48. Marso SP, Teirstein PS, Kereiakes DJ, Moses J, Lasala J, Grantham JA. Percutaneous coronary intervention use in the United States: defining measures of appropriateness. JACC Cardiovasc Interv. 2012 Feb;5(2):229-35.

49. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. J Clin Epidemiol. 2000 Dec;53(12):1258-67.

50. Shah BR, Cowper PA, O'Brien SM, Jensen N, Drawz M, Patel MR, et al. Patterns of cardiac stress testing after revascularization in community practice. J Am Coll Cardiol. 2010 Oct 12;56(16):1328-34.

51. Eisenberg MJ, Schechter D, Lefkovits J, Goudreau E, Deligonul U, Mak KH, et al. Use of routine functional testing after percutaneous transluminal coronary angioplasty: results from the ROSETTA Registry. Am Heart J. 2001 May;141(5):837-46.

52. Gidengil C, Garzon P, Eisenberg MJ. Functional testing after percutaneous transluminal coronary angioplasty in Canada and the United States: a survey of practice patterns. Can J Cardiol. 2000 Jun;16(6):739-46.

53. Harb SC, Cook T, Jaber WA, Marwick TH. Exercise testing in asymptomatic patients after revascularization: Are outcomes altered? Arch Intern Med. 2012;172(11):854-61.

54. Eisenberg MJ, Blankenship JC, Huynh T, Azrin M, Pathan A, Sedlis S, et al. Evaluation of routine functional testing after percutaneous coronary intervention. Am J Cardiol. 2004 Mar 15;93(6):744-7.

55. Eisenberg MJ, Wilson B, Lauzon C, Huynh T, Eisenhauer M, Mak KH, et al. Routine functional testing after percutaneous coronary intervention: results of the aggressive diagnosis of restenosis in high-risk patients (ADORE II) trial. Acta Cardiol. 2007 Apr;62(2):143-50.

56. San Roman JA, Vilacosta I, Castillo JA, Rollan MJ, Hernandez M, Peral V, et al. Selection of the optimal stress test for the diagnosis of coronary artery disease. Heart. 1998 Oct;80(4):370-6.

57. Schinkel AF, Bax JJ, Geleijnse ML, Boersma E, Elhendy A, Roelandt JR, et al. Noninvasive evaluation of ischaemic heart disease: myocardial perfusion imaging or stress echocardiography? European heart journal. 2003 May;24(9):789-800.

58. Kim C, Kwok YS, Heagerty P, Redberg R. Pharmacologic stress testing for coronary disease diagnosis: A meta-analysis. Am Heart J. 2001 Dec;142(6):934-44.

59. Gaur A, Yeon SB, Lewis CW, Manning WJ. Valvular flow abnormalities are often identified by a resting focused Doppler examination performed at the time of stress echocardiography. Am J Med. 2003 Jan;114(1):20-4.

60. Centers for Medicare and Medicaid Services. Physician Fee Schedule. Washington, DC 2011.

61. Navare SM, Mather JF, Shaw LJ, Fowler MS, Heller GV. Comparison of risk stratification with pharmacologic and exercise stress myocardial perfusion imaging: a meta-analysis. J Nucl Cardiol. 2004 Sep-Oct;11(5):551-61.

62. Goraya TY, Jacobsen SJ, Pellikka PA, Miller TD, Khan A, Weston SA, et al. Prognostic value of treadmill exercise testing in elderly persons. Ann Intern Med. 2000 Jun 6;132(11):862-70.

63. Federspiel JJ, Mudrick DW, Shah BR, Stearns SC, Masoudi FA, Cowper PA, et al. Patterns and predictors of stress testing modality after percutaneous coronary stenting: Data from the NCDR((R)). JACC Cardiovasc Imaging. 2012 Oct;5(10):969-80.

64. Levine GN, Bates ER, Blankenship JC, Bailey SR, Bittl JA, Cercek B, et al. 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention. A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. J Am Coll Cardiol. 2011 Dec 6;58(24):e44-122.

65. Rossi JS, Federspiel JJ, Crespin DJ, Carey TS, Sheridan BC, Stearns SC. Stress imaging utilization and repeat revascularization among Medicare patients with high risk coronary artery disease. Am J Cardiol. 2012 Jul 21;110(9):1270-4.

66. Brindis RG, Fitzgerald S, Anderson HV, Shaw RE, Weintraub WS, Williams JF. The American College of Cardiology-National Cardiovascular Data Registry (ACC-NCDR): building a national clinical data repository. J Am Coll Cardiol. 2001 Jun 15;37(8):2240-5.

67. National Cardiovascular Data Registry. CathPCI registry: program overview 2009 [July 31, 2011]. Available from: http://www.ncdr.com/webncdr/ncdrdocuments/CathPCI_Program_Overview.pdf.

68. Hammill BG, Hernandez AF, Peterson ED, Fonarow GC, Schulman KA, Curtis LH. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. Am Heart J. 2009 Jun;157(6):995-1000.

69. Douglas PS, Brennan JM, Anstrom KJ, Sedrakyan A, Eisenstein EL, Haque G, et al. Clinical effectiveness of coronary stents in elderly persons: results from 262,700 Medicare patients in the American College of Cardiology-National Cardiovascular Data Registry. J Am Coll Cardiol. 2009 May 5;53(18):1629-41.

70. Brennan JM, Peterson ED, Messenger JC, Rumsfeld JS, Weintraub WS, Anstrom KJ, et al. Linking the National Cardiovascular Data Registry CathPCI Registry with Medicare claims data: validation of a longitudinal cohort of elderly patients undergoing cardiac catheterization. Circ Cardiovasc Qual Outcomes. 2012 Jan;5(1):134-40.

71. Lin GA, Dudley RA, Lucas FL, Malenka DJ, Vittinghoff E, Redberg RF. Frequency of stress testing to document ischemia prior to elective percutaneous coronary intervention. JAMA. 2008 Oct 15;300(15):1765-73.

72. Dauerman HL. Reasonable incomplete revascularization. Circulation. 2011 May 31;123(21):2337-40.

73. Muller HG, Wang JL. Hazard rate estimation under random censoring with varying kernels and bandwidths. Biometrics. 1994 Mar;50(1):61-76.

74. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc. 1989;84(408):1065-73.

75. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat. 1988 Sep;16(3):1141-54.

76. Xiao Y, Abrahamowicz M. Bootstrap-based methods for estimating standard errors in Cox's regression analyses of clustered event times. Stat Med. 2010 Mar 30;29(7-8):915-23.

77. Venkitachalam L, Kip KE, Mulukutla SR, Selzer F, Laskey W, Slater J, et al. Temporal trends in patient-reported angina at 1 year after percutaneous coronary revascularization in the stent era: a report from the National Heart, Lung, and Blood Institute-sponsored 1997-2006 dynamic registry. Circ Cardiovasc Qual Outcomes. 2009 Nov;2(6):607-15.

78. Welch H, Hayes KJ, Frost C. Repeat testing among Medicare beneficiaries. Arch Intern Med. 2012:1-7.

79. Lucas FL, Sirovich BE, Gallagher PM, Siewers AE, Wennberg DE. Variation in cardiologists' propensity to test and treat: is it associated with regional variation in utilization? Circ Cardiovasc Qual Outcomes. 2010 May;3(3):253-60.

80. Shreibati JB, Baker LC, Hlatky MA. Association of coronary CT angiography or stress testing with subsequent utilization and spending among Medicare beneficiaries. JAMA. 2011 Nov 16;306(19):2128-36.

81. Shaw LJ, Mieres JH, Hendel RH, Boden WE, Gulati M, Veledar E, et al. Comparative effectiveness of exercise electrocardiography with or without myocardial perfusion single photon emission computed tomography in women with suspected coronary artery disease: results from the What Is the Optimal Method for Ischemia Evaluation in Women (WOMEN) trial. Circulation. 2011 Sep 13;124(11):1239-49.

82. Cutler DM, Ghosh K. The potential for cost savings through bundled episode payments. N Engl J Med. 2012 Mar 22;366(12):1075-7.

83. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998 Jan;36(1):8-27.

84. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol. 2006 Jun 15;163(12):1149-56.

85. Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: an application to data on right heart catheteriation. Health Serv Outcomes Res Methodol. 2001;2:259-78.

86. Austin PC. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. Stat Med. 2010 Sep 10;29(20):2137-48.

87. Austin PC. Primer on statistical interpretation or methods report card on propensity-score matching in the cardiology literature from 2004 to 2006: a systematic review. Circ Cardiovasc Qual Outcomes. 2008 Sep;1(1):62-7.

88. Bang H, Tsiatis AA. Estimating medical costs with censored data. Biometrika. 2000;87(2):329-43.

89. Shaw LJ, Marwick TH, Berman DS, Sawada S, Heller GV, Vasey C, et al. Incremental cost-effectiveness of exercise echocardiography vs. SPECT imaging for the evaluation of stable chest pain. European heart journal. 2006 Oct;27(20):2448-58.

90. Bart BA, Erlien DA, Herzog CA, Asinger RW. Marked differences between patients referred for stress echocardiography and myocardial perfusion imaging studies. Am Heart J. 2005 May;149(5):888-93.

91. Heijenbrok-Kal MH, Fleischmann KE, Hunink MG. Stress echocardiography, stress single-photon-emission computed tomography and electron beam computed tomography for the assessment of coronary artery disease: a meta-analysis of diagnostic performance. Am Heart J. 2007 Sep;154(3):415-23.

92. Adriaensen ME, Kock MC, Stijnen T, van Sambeek MR, van Urk H, Pattynama PM, et al. Peripheral arterial disease: therapeutic confidence of CT versus digital subtraction angiography and effects on additional imaging recommendations. Radiology. 2004 Nov;233(2):385-91.

93. Hachamovitch R, Hayes S, Friedman JD, Cohen I, Shaw LJ, Germano G, et al. Determinants of risk and its temporal variation in patients with normal stress myocardial perfusion scans: what is the warranty period of a normal scan? Journal of the American College of Cardiology. 2003 Apr 16;41(8):1329-40.

94. Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med. 2007 Apr 12;356(15):1503-16.

95. Lauer MS. Comparative effectiveness research: the view from the NHLBI. J Am Coll Cardiol. 2009 Mar 24;53(12):1084-6.

96. Drozda JP, Bufalino VJ, Fasules JW, Peterson ED, Redberg RF, Wright JS, et al. ACC 2009 Advocacy Position Statement: Principles for comparative effectiveness research. J Am Coll Cardiol. 2009 Oct 27;54(18):1744-6.

97. Gunasekara FI, Carter K, Blakely T. Glossary for econometrics and epidemiology. J Epidemiol Community Health. 2008 Oct;62(10):858-61.

98. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006 Mar;98(3):253-9.

99. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol. 2003 Aug 1;158(3):280-7.

100. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics. 2005 Dec;61(4):962-73.

101. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol. 2006 May;59(5):437-47.

102. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. JAMA. 1994 Sep 21;272(11):859-66.

103. Fang G, Brooks JM, Chrischilles EA. A new method to isolate local-area practice styles in prescription use as the basis for instrumental variables in comparative effectiveness research. Med Care. 2010 Aug;48(8):710-7.

104. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. Int J Biostat. 2007;3(1):14.

105. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health. 1998;19:17-34.

106. Harris KM, Remler DK. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. Health services research. 1998 Dec;33(5 Pt 1):1337-60.

107. Stolker JM, Patil H, Chhatriwalla AK, Venkitachalam L, Kennedy KF, Jones PG, et al. Temporal Trends of Drug Eluting Stent Use After Myocardial Infarction: Does More Selective Utilization Impact Health Status Outcomes? [Conference Presentation]. Quality of Care and Outcomes Research in Cardiovascular Disease and Stroke Scientific Sessions; Washington 2010.

108. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. Econometrica. 1994;62(2):467-75.

109. Heckman JJ, Urzua S, Vytlacil E. Understanding instrumental variables in models with essential heterogeneity. Rev Econ Stat. 2006;88(3):389-432.

110. Carneiro P, Heckman JJ, Vytlacil E. Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. Econometrica. 2010 Jan 1;78(1):377-94.

111. Heckman JJ, Vytlacil E. Structural equations, treatment effects, and econometric policy evaluation. Econometrica. 2005;73(3):669-738.

112. Heckman JJ, Vytlacil EJ. Local instrumental variables and latent variable models for identifying and bounding treatment effects. Proceedings of the National Academy of Sciences of the United States of America. 1999 Apr 13;96(8):4730-4.

113. Basu A, Heckman JJ, Navarro-Lozano S, Urzua S. Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. Health Econ. 2007 Nov;16(11):1133-57.

114. Basu A. Estimating decision-relevant comparative effects using instrumental variables. Stat Biosci. 2011;3(1):6-27.

115. Heckman JJ, Vytlacil EJ. Chapter 70: Economic evaluation of social programs. Part I: Causal models, structural models and econometric policy evaluation. In: Heckman JJ, Leamer EE, editors. Handbook of econometrics. 6B: Elsevier B.V.; 2007. p. 4779-874.

116. Heckman JJ, Vytlacil EJ. Chapter 71: Econometric evaluation of social programs, Part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In: Heckman JJ, Leamer EE, editors. Handbook of econometrics. 6B: Elsevier B.V.; 2007. p. 4875-5143.

117. Abbring JH, Heckman JJ. Chapter 72: Econometric evaluation of social programs, Part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In: Heckman JJ, Leamer EE, editors. Handbook of econometrics. 6B: Elsevier B.V.; 2007. p. 5145-303.

118. Heidenreich PA. Measuring value from the patient's perspective. Circ Cardiovasc Qual Outcomes. 2011 Jan 1;4(1):9-10.

119. Deshpande NV, Serruys PW. Current status of plain old balloon angioplasty. Indian Heart J. 1998 Oct;50 Suppl 1:5-13.

120. Kimmel SE, Localio AR, Brensinger C, Miles C, Hirshfeld J, Haber HL, et al. Effects of coronary stents on cardiovascular outcomes in broad-based clinical practice. Archives of internal medicine. 2000 Sep 25;160(17):2593-9.

121. Kimmel SE, Localio AR, Krone RJ, Laskey WK. The effects of contemporary use of coronary stents on in-hospital mortality. Registry Committee of the Society for Cardiac Angiography and Interventions. J Am Coll Cardiol. 2001 Feb;37(2):499-504.

122. Fischman DL, Leon MB, Baim DS, Schatz RA, Savage MP, Penn I, et al. A randomized comparison of coronary-stent placement and balloon angioplasty in the treatment of coronary artery disease. Stent Restenosis Study Investigators. N Engl J Med. 1994 Aug 25;331(8):496-501.

123. Serruys PW, de Jaegere P, Kiemeneij F, Macaya C, Rutsch W, Heyndrickx G, et al. A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. Benestent Study Group. N Engl J Med. 1994 Aug 25;331(8):489-95.

124. Sutton JM, Ellis SG, Roubin GS, Pinkerton CA, King SB, 3rd, Raizner AE, et al. Major clinical events after coronary stenting. The multicenter registry of acute and elective Gianturco-Roubin stent placement. The Gianturco-Roubin Intracoronary Stent Investigator Group. Circulation. 1994 Mar;89(3):1126-37.

125. Williams DO, Holubkov R, Yeh W, Bourassa MG, Al-Bassam M, Block PC, et al. Percutaneous coronary intervention in the current era compared with 1985-1986: the National Heart, Lung, and Blood Institute Registries. Circulation. 2000 Dec 12;102(24):2945-51.

126. Haude M, Konorza TF, Kalnins U, Erglis A, Saunamaki K, Glogar HD, et al. Heparin-coated stent placement for the treatment of stenoses in small coronary arteries of symptomatic patients. Circulation. 2003 Mar 11;107(9):1265-70.

127. Teirstein PS, Massullo V, Jani S, Popma JJ, Russo RJ, Schatz RA, et al. Three-year clinical and angiographic follow-up after intracoronary radiation : results of a randomized clinical trial. Circulation. 2000 Feb 1;101(4):360-5.

128. Colombo A, Drzewiecki J, Banning A, Grube E, Hauptmann K, Silber S, et al. Randomized study to assess the effectiveness of slow- and moderate-release polymer-based paclitaxel-eluting stents for coronary artery lesions. Circulation. 2003 Aug 19;108(7):788-94.

129. Morice MC, Serruys PW, Sousa JE, Fajadet J, Ban Hayashi E, Perin M, et al. A randomized comparison of a sirolimus-eluting stent with a standard stent for coronary revascularization. N Engl J Med. 2002 Jun 6;346(23):1773-80.

130. Grube E, Silber S, Hauptmann KE, Mueller R, Buellesfeld L, Gerckens U, et al. TAXUS I: six- and twelve-month results from a randomized, double-blind trial on a slow-release paclitaxel-eluting stent for de novo coronary lesions. Circulation. 2003 Jan 7;107(1):38-42.

131. Rao SV, Shaw RE, Brindis RG, Klein LW, Weintraub WS, Peterson ED. On- versus off-label use of drug-eluting coronary stents in clinical practice (report from the American College of Cardiology National Cardiovascular Data Registry [NCDR]). Am J Cardiol. 2006 May 15;97(10):1478-81.

132. Jeremias A, Sylvia B, Bridges J, Kirtane AJ, Bigelow B, Pinto DS, et al. Stent thrombosis after successful sirolimus-eluting stent implantation. Circulation. 2004 Apr 27;109(16):1930-2.

133. Mauri L, Hsieh WH, Massaro JM, Ho KK, D'Agostino R, Cutlip DE. Stent thrombosis in randomized clinical trials of drug-eluting stents. N Engl J Med. 2007 Mar 8;356(10):1020-9.

134. Pfisterer M, Brunner-La Rocca HP, Buser PT, Rickenbacher P, Hunziker P, Mueller C, et al. Late clinical events after clopidogrel discontinuation may limit the benefit of drug-eluting stents: an observational study of drug-eluting versus bare-metal stents. Journal of the American College of Cardiology. 2006 Dec 19;48(12):2584-91.

135. Nordmann AJ, Briel M, Bucher HC. Mortality in randomized controlled trials comparing drug-eluting vs. bare metal stents in coronary artery disease: a meta-analysis. Eur Heart J. 2006 Dec;27(23):2784-814.

136. Wenaweser P, Daemen J, Zwahlen M, van Domburg R, Juni P, Vaina S, et al. Incidence and correlates of drug-eluting stent thrombosis in routine clinical practice. 4-year results from a large 2-institutional cohort study. Journal of the American College of Cardiology. 2008 Sep 30;52(14):1134-40.

137. Spertus JA, Kettelkamp R, Vance C, Decker C, Jones PG, Rumsfeld JS, et al. Prevalence, predictors, and outcomes of premature discontinuation of thienopyridine therapy after drug-eluting stent placement: results from the PREMIER registry. Circulation. 2006 Jun 20;113(24):2803-9.

138. Roe MT, Chen AY, Cannon CP, Rao S, Rumsfeld J, Magid DJ, et al. Temporal changes in the use of drug-eluting stents for patients with non-ST-segment-elevation myocardial infarction undergoing percutaneous coronary intervention from 2006 to 2008: results from the can rapid risk stratification of unstable angina patients supress adverse outcomes with early implementation of the ACC/AHA guidelines (CRUSADE) and acute coronary treatment and intervention outcomes network-get with the guidelines (ACTION-GWTG) registries. Circ Cardiovasc Qual Outcomes. 2009 July 28, 2009:CIRCOUTCOMES.109.850248.

139. Grines CL, Bonow RO, Casey DE, Jr., Gardner TJ, Lockhart PB, Moliterno DJ, et al. Prevention of premature discontinuation of dual antiplatelet therapy in patients with coronary

artery stents: a science advisory from the American Heart Association, American College of Cardiology, Society for Cardiovascular Angiography and Interventions, American College of Surgeons, and American Dental Association, with representation from the American College of Physicians. Journal of the American College of Cardiology. 2007 Feb 13;49(6):734-9.

140. Tu JV, Bowen J, Chiu M, Ko DT, Austin PC, He Y, et al. Effectiveness and safety of drug-eluting stents in Ontario. N Engl J Med. 2007 Oct 4;357(14):1393-402.

141. Mauri L, Silbaugh TS, Garg P, Wolf RE, Zelevinsky K, Lovett A, et al. Drug-eluting or bare-metal stents for acute myocardial infarction. N Engl J Med. 2008 Sep 25;359(13):1330-42.

142. Applegate RJ, Sacrinty MT, Kutcher MA, Santos RM, Gandhi SK, Baki TT, et al. "Off-label" stent therapy 2-year comparison of drug-eluting versus bare-metal stents. J Am Coll Cardiol. 2008 Feb 12;51(6):607-14.

143. Brodie BR, Wilson H, Stuckey T, Nussbaum M, Laurent S, Bradshaw B, et al. Outcomes with drug-eluting versus bare-metal stents in saphenous vein graft intervention results from the STENT (strategic transcatheter evaluation of new therapies) group. JACC Cardiovasc Interv. 2009 Nov;2(11):1105-12.

144. Rao SV, Shaw RE, Brindis RG, Klein LW, Weintraub WS, Krone RJ, et al. Patterns and outcomes of drug-eluting coronary stent use in clinical practice. Am Heart J. 2006 Aug;152(2):321-6.

145. Malenka DJ, Kaplan AV, Lucas FL, Sharp SM, Skinner JS. Outcomes following coronary stenting in the era of bare-metal vs the era of drug-eluting stents. JAMA. 2008 Jun 25;299(24):2868-76.

146. Hannan EL, Racz M, Holmes DR, Walford G, Sharma S, Katz S, et al. Comparison of coronary artery stenting outcomes in the eras before and after the introduction of drug-eluting stents. Circulation. 2008 Apr 22;117(16):2071-8.

147. Federspiel JJ, Stearns SC, Sheridan BC, Kuritzky JJ, D'Arcy LP, Crespin DJ, et al. Evaluating the effectiveness of a rapidly adopted cardiovascular technology with administrative data: the case of drug-eluting stents for acute coronary syndromes. Am Heart J. 2012 Aug;164(2):207-14.

148. Yeh RW, Chandra M, McCulloch CE, Go AS. Accounting for the mortality benefit of drug-eluting stents in percutaneous coronary intervention: a comparison of methods in a retrospective cohort study. BMC Med. 2011;9:78.

149. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. Am J Epidemiol. 2012 Jan 1;175(1):60-5.

150. Marshall DA. Principles of economics. New York: Macmillan; 1890.

151. Roy AD. Some thoughts on the distribution of earnings. Oxf Econ Pap. 1951;3(2):135-46.

152. Neyman J, Iwaszkiewicz K, Kolodziejczyk S. Statistical problems in agricultural experimentation. J R Stat Soc. 1935;Suppl. 2(2):107-80.

153. Quandt RE. The estimation of the parameters of a linear regression system obeying two separate regimes. J Am Stat Assoc. 1958;53(284):873-80.

154. Heckman J. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. J Hum Res. 1997;32(3):441-62.

155. Angrist JD, Imbens GW. Comment on James J. Heckman, "Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations". J Hum Res. 1999;34(4):823-7.

156. Heckman JJ. Instrumental variables: Response to Angrist and Imbens. J Hum Res. 1999;34(4):828-37.

157. Heckman JJ, Urzúa S. Comparing IV with structural models: What simple IV can and cannot identify. Journal of Econometrics. 2010;156(1):27-37.

158. Deaton A. Instruments in development: randomization in the tropics, and the search for the elusive keys to economic development. Proc Brit Academy. 2009;162:123-60.

159. Imbens GW. Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). Journal of Economic Literature. 2010;48(2):399-423.

160. Curtis LH, Hammill BG, Eisenstein EL, Kramer JM, Anstrom KJ. Using inverse probability-weighted estimators in comparative effectiveness analyses with observational databases. Med Care. 2007 Oct;45(10 Supl 2):S103-7.

161. Angrist JD. Estimating the labor market impact of voluntary military service using social security date on military applicants. Econometrica. 1998;66:249-88.

162. Björklund A, Moffitt R. The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. The Review of Economics and Statistics. 1987;69(1):42-9.

163. Basu A. Estimating person-centered treatment (PeT) effects using instrument variables. NBER Working Papers. 2012;W18056.

164. Tsai TT, Messenger JC, Brennan JM, Patel UD, Dai D, Piana RN, et al. Safety and Efficacy of Drug-Eluting Stents in Older Patients With Chronic Kidney Disease. Journal of the American College of Cardiology. 2011 October 25, 2011;58(18):1859-69.

165. Westreich D, Cole SR, Funk MJ, Brookhart MA, Sturmer T. The role of the c-statistic in variable selection for propensity score models. Pharmacoepidemiol Drug Saf. 2011 Mar;20(3):317-20.

166. Efron B. An introduction to the bootstrap. Tibshirani R, editor. Boca Raton: Chapman & Hall; 1993.

167. Haukoos JS, Lewis RJ. Advanced statistics: bootstrapping confidence intervals for statistics with "difficult" distributions. Acad Emerg Med. 2005 Apr;12(4):360-5.

168. Geronimus AT, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. Am J Epidemiol. 1998 Sep 1;148(5):475-86.

169. Poses RM, Smith WR, McClish DK, Anthony M. Controlling for confounding by indication for treatment. Are administrative data equivalent to clinical data? Med Care. 1995 Apr;33(4 Suppl):AS36-46.

170. Hammill BG, Curtis LH, Fonarow GC, Heidenreich PA, Yancy CW, Peterson ED, et al. Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. Circ Cardiovasc Qual Outcomes. 2011 Jan 1;4(1):60-7.

171. Cooper AL, Trivedi AN. Fitness memberships and favorable selection in Medicare Advantage plans. N Engl J Med. 2012 Jan 12;366(2):150-7.

172. Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. Am J Epidemiol. 2011 Mar 1;173(5):569-77.

173. Chandra A, Staiger DO. Productivity spillovers in healthcare: Evidence from the treatment of heart attacks. J Polit Econ. 2007;115:103-40.

174. Chin CT, Wang TY, Anstrom KJ, Zhu B, Maa JF, Messenger JC, et al. Treatment with adenosine diphosphate receptor inhibitors-longitudinal assessment of treatment patterns and events after acute coronary syndrome (TRANSLATE-ACS) study design: expanding the paradigm of longitudinal observational research. Am Heart J. 2011 Nov;162(5):844-51.

175. Blue Cross Blue Shield of Delaware. Statement regarding DOI/BCBSD settlement agreement 2011.

176. Eisenstein EL, Anstrom KJ, Kong DF, Shaw LK, Tuttle RH, Mark DB, et al. Clopidogrel use and long-term clinical outcomes after drug-eluting stent implantation. JAMA. 2007 Jan 10;297(2):159-68.